| Europäisches Patentamt | European Patent Office | Office européen des brevets |

Bescheinigung   Certificate   Attestation

| Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein. | The attached documents are exact copies of the European patent application described on the following page, as originally filed. | Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante. |

Patentanmeldung Nr.   Patent application No.   Demande de brevet n°

04090187.8

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

Anmeldung Nr:
Application no.: 04090187.8
Demande no:

Anmeldetag:
Date of filing: 10.05.04
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Epigenomics AG
Kastanienallee 24
10435 Berlin
ALLEMAGNE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Methods and nucleic acids for the analysis of CpG dinucleotide methylation status
associated with the development of prostate cancer

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s)
revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

EP/10.02.04/EP 04090040
EP/01.12.03/EP 03090414

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

C12Q1/68

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of
filing/Etats contractants désignées lors du dépôt:

   AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL
   PL PT RO SE SI SK TR LI

METHODS AND NUCLEIC ACIDS FOR THE ANALYSIS OF CpG DINUCLEOTIDE METHYLATION STATUS ASSOCIATED WITH THE DEVELOPMENT OF PROSTATE CANCER.

FIELD OF THE INVENTION

The present invention relates to human DNA sequences that exhibit altered methylation patterns (hypermethylation or hypomethylation) in cancer patients. Particular embodiments of the invention provide highly accurate methods for detection and differentiation of prostate carcinomas.

BACKGROUND

Correlation of aberrant DNA methylation with cancer. Aberrant DNA methylation within CpG 'islands' is characterized by *hyper-* or *hypomethylation* of CpG dinucleotide sequences leading to abrogation or overexpression of a broad spectrum of genes, and is among the earliest and most common alterations found in, and correlated with human malignancies. Additionally, abnormal methylation has been shown to occur in CpG-rich regulatory elements in intronic and coding parts of genes for certain tumors. In colon cancer, aberrant DNA methylation constitutes one of the most prominent alterations and inactivates many tumor suppressor genes including, *inter alia*, p14ARF, p16INK4a, THBS1, MINT2, and MINT31 and DNA mismatch repair genes such as hMLH1.

Aside from the specific hypermethylation of tumor suppressor genes, an overall hypomethylation of DNA can be observed in tumor cells. This decrease in global methylation can be detected early, far before the development of frank tumor formation. A correlation between hypomethylation and increased gene expression has been determined for many oncogenes.

Prostate cancer. The prostate is a male sex accessory gland, comprising about 30 to 50 branched glands. It is surrounded by a fibroelastic capsule that separates the gland into discrete lobes. The *central* zone of the organ is composed of pseudo stratified epithelium, the *peripheral* zone comprises the bulk of the organ and the two tissue types are separated by a *transitional* zone.

Benign prostate hypertrophy is present in about 50% of men aged 50 or above, and in 95% of men aged 75 or above. Prostate

cancer is a significant health care problem in Western countries with an incidence of 180 per 100,000 in the United States in 1999 (*Cancer J. Clin.*, 49:8, 1999).

Diagnosis and prognosis of prostate cancer; deficiencies of prior art approaches. Different screening strategies have been employed with at least some degree of success to improve early detection of prostate cancer, including determination of levels of prostate specific antigen ("PSA") and digital rectal examination. If a prostate carcinoma is suspected in a patient, diagnosis of cancer is confirmed or excluded by the histological and cytological analysis of biopsy samples for features associated with malignant transformation. The *zone of origin* of a prostatic cell proliferative disorder is currently determined by the 'PSA density.' PSA density is determined by dividing the weight of the prostate (as estimated by transrectal ultrasound) by the prostate specific antigen levels of the patient. Levels of over 15% percent are considered as indicative of prostate cancer and grounds for a biopsy. The biopsy, in turn, is used for histological and cytological analysis to determine the *zone of origin*.

However, using routine histological examination, it is often difficult to distinguish benign hyperplasia of the prostate from early stages of prostate carcinoma, even if an adequate biopsy is obtained (McNeal J. E. et al., *Hum. Pathol.* 2001, 32:441-6). Furthermore, small or otherwise insufficient biopsy samples often impede the analysis.

Molecular markers would offer the advantage that they could be used to efficiently analyze even very small tissue samples, and samples whose tissue architecture has not been maintained. Within the last decade, numerous genes have been studied with respect to differential expression among benign hyperplasia of the prostate and different grades of prostate cancer.

However, no single marker has as yet been shown to be sufficient for the diagnosis of prostate tumors in a clinical setting.

Alternatively, high-dimensional mRNA based approaches may, in particular instances, provide a means to distinguish between different tumor types and benign and malignant lesions. However, application of such approaches as a routine diagnostic

tool in a clinical environment is impeded and substantially limited by the extreme instability of mRNA, the rapidly occurring expression changes following certain triggers (*e.g.*, sample collection), and, most importantly, by the large amount of mRNA needed for analysis which often cannot be obtained from a routine biopsy (see, *e.g.*, Lipshutz, R. J. et al., *Nature Genetics* 21:20-24, 1999; Bowtell, D. D. L. *Nature Genetics Suppl.* 21:25-32, 1999).

The GSTP1 gene. The core promoter region of the Gluthione S-Transferase P gene (GSTP1; accession no. NM_000852) has been shown to be hypermethylated in prostate tumor tissue. The glutathione S-transferase pi enzyme is involved in the detoxification of electrophilic carcinogens, and impaired or decreased levels of enzymatic activity (GSTPi impairment) have been associated with the development of neoplasms, particularly in the prostate. Mechanisms of GSTPi impairment include mutation (the GSTP*B allele has been associated with a higher risk of cancer) and methylation.

Prior art GSTP1 studies. Lee et al., in United States Patent No 5,552,277, disclosed that the expression of the gluthione-S-transferase (GST) Pi gene was downregulated in a significant proportion of prostate carcinomas. Moreover, by means of restriction enzyme analysis they were able to show that the promoter region of the of the GSTPi gene was upmethylated (hypermethylated) in prostate carcinomas as opposed to normal prostate and leukocyte tissue. However, due to the limited and imprecise nature of the analysis technique used (HpaIII digestion, followed by Southern blotting) the exact number and position of the methylated CG dinucleotides were not characterized.

Douglas et al. (WO9955905) used a method comprising bisulfite treatment, followed by methylation specific PCR to show that prostate carcinoma-specific GSTPi hypermethylation was localized to the core promoter regions, and localized a number of CpG positions that had not been characterised by Lee et al.

Herman and Baylin (United States Patent No. 6,017,704) describe the use of methylation specific primers for methylation analysis, and describe a particular primer pair

suitable for the analysis of the corresponding methylated GSTPi promoter sequence.

However, with respect to the use of GSTPi markers, the prior art is limited with respect to the number of GSTPi promoter CpG sequences that have been characterized for differential methylation status. Moreover, there are no disclosures, suggestions or teachings in the prior art of how such markers could be used to distinguish among benign hyperplasia of the prostate and different grades of prostate cancer.

Aberrant genetic methylation has also been observed in several other genes including AR, p16 (CDKN2a/INK4a), CD44, CDH1. Genome wide hypomethylation for example of the LINE-1 repetitive element has also been associated with tumor progression (Santourlidis S ,Florl A ,Ackermann R ,Wirtz HC ,Schulz WA 'High frequency of alterations in DNA methylation in adenocarcinoma of the prostate.' Prostate 1999 May 15;39(3):166-74).

However, use of these genes as alternative or supplemental diagnostic, or otherwise clinically useful markers in a commercial setting has not been enabled. The application of differentially methylated genes to clinically utilizable platforms requires much further investigation into the sensitivity and specificity of the genes. For example, in the case of the gene CD44, a known metastasis suppressor, downregulation was associated with hypermethylation. However the use of this gene as a commercially avalable marker was not enabled as it was also methylated in normal tissues. See Vis AN Oomen M Schroder FH van der Kwast TH 'Feasibility of assessment of promoter methylation of the CD44 gene in serum of prostate cancer patients.' Mol Urol. 2001 Winter;5(4):199-203.

Development of medical tests. Two key evaluative measures of any medical screening or diagnostic test are its sensitivity and specificity, which measure how well the test performs to accurately detect all affected individuals without exception, and without falsely including individuals who do not have the target disease (predictive value). Historically, many diagnostic tests have been criticized due to poor sensitivity and specificity.

A true positive (TP) result is where the test is positive and the condition is present. A false positive (FP) result is where the test is positive but the condition is not present. A true negative (TN) result is where the test is negative and the condition is not present. A false negative (FN) result is where the test is negative but the condition is not present.

Sensitivity = TP/(TP+FN)
Specificity = TN/(FP+TN)
Predictive value = TP/(TP+FP)

Sensitivity is a measure of a test's ability to correctly detect the target disease in an individual being tested. A test having poor sensitivity produces a high rate of false negatives, i.e., individuals who have the disease but are falsely identified as being free of that particular disease. The potential danger of a false negative is that the diseased individual will remain undiagnosed and untreated for some period of time, during which the disease may progress to a later stage wherein treatments, if any, may be less effective. An example of a test that has low sensitivity is a protein-based blood test for HIV. This type of test exhibits poor sensitivity because it fails to detect the presence of the virus until the disease is well established and the virus has invaded the bloodstream in substantial numbers. In contrast, an example of a test that has high sensitivity is viral-load detection using the polymerase chain reaction (PCR). High sensitivity is achieved because this type of test can detect very small quantities of the virus. High sensitivity is particularly important when the consequences of missing a diagnosis are high.

Specificity, on the other hand, is a measure of a test's ability to identify accurately patients who are free of the disease state. A test having poor specificity produces a high rate of false positives, i.e., individuals who are falsely identified as having the disease. A drawback of false positives is that they force patients to undergo unnecessary medical procedures treatments with their attendant risks, emotional and

financial stresses, and which could have adverse effects on the patient's health. A feature of diseases which makes it difficult to develop diagnostic tests with high specificity is that disease mechanisms, particularly in cancer, often involve a plurality of genes and proteins. Additionally, certain proteins may be elevated for reasons unrelated to a disease state. An example of a test that has high specificity is a gene-based test that can detect a p53 mutation. Specificity is important when the cost or risk associated with further diagnostic procedures or further medical intervention are very high.

The PSA blood test has a sensitivity of 73%, specificity of 60% and predictive value of 31.5%. PSA sensitivity and specificity can be improved but involve tradeoffs. PSA sensitivity can be improved by adjusting the "normal" PSA level to a lower value for younger men or by following serum PSA values in an individual patient over time (PSA velocity). Both methods will increase the number of cancers detected, but they also increase the number of men undergoing biopsy. Conversely, specificity can be improved by using higher "normal" PSA levels for older men, by using the free-to-total PSA ratio, or by adjusting the normal value according to the size of the prostate. These three methods decrease the number of unnecessary biopsies, but they increase the risk that some cancers will be missed.
It can therefore be seen that there exists a need for a means of prostate cancer diagnosis with improved sensitivity, specificity and/or predictive value.

Sensitivity and specificity of quantitative methylation-specific polymerase chain reaction (QMSP) assay alone (without histological analysis) in prostate cancer analysis of needle biopsies has ranged from 30% sensitivity and 100% specificity to 89% sensitivity and 64% specificity (Harden *et. al.* J Natl Cancer Inst 2003; 95: 1634-1637). However the predictive value of said technique as a clinical screening tool was not analysed. Furthermore, genetic testing of serum and bodily fluids such as urine and saliva would reduce the need for biopsies to detect cancer and would thus be the most effective screening or monitering tool. However the development of such tests requires an extremely high degree of sensitivity

and specificity. Analysis of GSTPi gene hypermethylation (Cairns P, Esteller M, Herman JG, Schoenberg M, Jeronimo C, Sanchez-Cespedes M, et al. Molecular detection of prostate cancer in urine by GSTP1 hypermethylation. Clin Cancer Res 2001;7:2727-30.) in urine sediment of prostate cancer patients showed that only 6 out of 22 individuals with elevated methylation levels in biopsied tumors showed corresponding hypermethylation in urine samples.

Multifactorial approach. Cancer diagnostics has traditionally relied upon the detection of single molecular markers (e.g. gene mutations, elevated PSA levels). Unfortunately, cancer is a disease state in which single markers have typically failed to detect or differentiate many forms of the disease. Thus, assays that recognize only a single marker have been shown to be of limited predictive value. A fundamental aspect of this invention is that methylation based cancer diagnostics and the screening, diagnosis, and therapeutic monitoring of such diseases will provide significant improvements over the state-of-the-art that uses single marker analyses by the use of a selection of multiple markers. The multiplexed analytical approach is particularly well suited for cancer diagnostics since cancer is not a simple disease, this multi-factorial "panel" approach is consistent with the heterogeneous nature of cancer, both cytologically and clinically.

Key to the successful implementation of a panel approach to methylation based diagnostic tests is the design and development of optimized panels of markers that can characterize and distinguish disease states. This patent application describes an efficient and unique panel of genes the methylation analysis of one or a combination of the members of the panel enabling the detection of cell proliferative disorders of the prostate with a particularly high sensitivity, specificity and/or predictive value.

Pronounced need in the art. Therefore, in view of the incidence of prostate hyperplasia (50% of men aged 50 or above, and 95% of men aged 75 or above) and prostate cancer (180 per 100,000), there is a substantial need in the art for the development of molecular markers that could be used to

effectively distinguish among benign hyperplasia of the prostate and different grades of prostate cancer. Additionally, there is a pronounced need in the art for the development of molecular markers that could be used to provide sensitive, accurate and non-invasive methods (as opposed to, e.g., biopsy and transrectal ultrasound) for the diagnosis, prognosis and treatment of prostate cell proliferative disorders.

SUMMARY OF THE INVENTION

The disclosed invention provides a means for detection of or differentiation between prostate cell proliferative disorders by analysis of a gene panel, with a sensitivity and specificity suitable for use in a body fluid or serum assay. The present invention provides novel methods for detecting or distinguishing between prostate cell proliferative disorders with a sensitivity of greater than 30% and a specificity of greater than 65%. Said method is most preferably utilised for detecting or detecting and distinguishing between prostate cell proliferative disorders. The invention provides a method for the analysis of biological samples for features associated with the development of prostate cell proliferative disorders, the method characterised in that at least one nucleic acid, or a fragment thereof, from the group consisting of SEQ ID NO: 1 to SEQ ID NO: 30 is/are contacted with a reagent or series of reagents capable of distinguishing between methylated and non methylated CpG dinucleotides within the genomic sequence, or sequences of interest.

The present invention provides a method for ascertaining genetic and/or epigenetic parameters of genomic DNA. The method has utility for the improved diagnosis, treatment and monitoring of prostate cell proliferative disorders, more specifically by enabling the improved identification of and differentiation between subclasses of said disorder and the genetic predisposition to said disorders. The invention presents improvements over the state of the art in that it enables a more specific and sensitive classification of prostate cell proliferative disorders than that achieved by

currently used tests thereby allowing for improved and informed treatment of patients.

Preferably, the source of the test sample is selected from the group consisting of cells or cell lines, histological slides, biopsies, paraffin-embedded tissue, bodily fluids, ejaculate, urine, blood, and combinations thereof. Preferably, the source is biopsies, bodily fluids, ejaculate, urine, or blood.

Specifically, the present invention provides a method for detecting prostate cell proliferative disorders with a sensitivity of greater than 30% and a specificity of greater than 65%, comprising: obtaining a biological sample comprising genomic nucleic acid(s); contacting the nucleic acid(s), or a fragment thereof, with one reagent or a plurality of reagents sufficient for distinguishing between methylated and non methylated CpG dinucleotide sequences within a target sequence of the subject nucleic acid, wherein the target sequence comprises, or hybridizes under stringent conditions to, a sequence comprising at least 16 contiguous nucleotides of SEQ ID NO: 1 to 30, said contiguous nucleotides comprising at least one CpG dinucleotide sequence; and determining, based at least in part on said distinguishing, the methylation state of at least one target CpG dinucleotide sequence, or an average, or a value reflecting an average methylation state of a plurality of target CpG dinucleotide sequences. Preferably, distinguishing between methylated and non methylated CpG dinucleotide sequences within the target sequence comprises methylation state-dependent conversion or non-conversion of at least one such CpG dinucleotide sequence to the corresponding converted or non-converted dinucleotide sequence within a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 28, and contiguous regions thereof corresponding to the target sequence.

Additional embodiments provide a method for the detection of prostate cell proliferative disorders with a sensitivity of greater than 30% and a specificity of greater than 65%, comprising: obtaining a biological sample having subject genomic DNA; extracting the genomic DNA; treating the genomic DNA, or a fragment thereof, with one or more reagents to convert 5-position unmethylated cytosine bases to uracil or to another base that is detectably dissimilar to cytosine in terms

of hybridization properties; contacting the treated genomic DNA, or the treated fragment thereof, with an amplification enzyme and at least two primers comprising, in each case a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent. or stringent conditions to a sequence selected from the group consisting SEQ ID NO: 5 to SEQ ID NO: 28, and complements thereof, wherein the treated DNA or the fragment thereof is either amplified to produce an amplificate, or is not amplified; and determining, based on a presence or absence of, or on a property of said amplificate, the methylation state of at least one CpG dinucleotide sequence selected from the group consisting of SEQ ID NO: 1 to SEQ ID NO: 4 and SEQ ID NOs: 29 & 30, or an average, or a value reflecting an average methylation state of a plurality of CpG dinucleotide sequences thereof. Preferably, at least one such hybridizing nucleic acid molecule or peptide nucleic acid molecule is bound to a solid phase. Preferably, determining comprises use of at least two methods selected from the group consisting of: hybridizing at least one nucleic acid molecule comprising a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 28, and complements thereof; hybridizing at least one nucleic acid molecule, bound to a solid phase, comprising a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 28, and complements thereof; hybridizing at least one nucleic acid molecule comprising a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 28, and complements thereof, and extending at least one such hybridized nucleic acid molecule by at least one nucleotide base; and sequencing of the amplificate.

Additional embodiments provide novel genomic and chemically modified nucleic acid sequences, as well as oligonucleotides and/or PNA-oligomers for analysis of cytosine methylation

patterns within sequences from the group consisting of SEQ ID
NO: 1 to SEQ ID NO: 4 and SEQ ID NOs: 29 & 30.

BRIEF DESCRIPTION OF THE DRAWINGS

DETAILED DESCRIPTION OF THE INVENTION
Definitions:
The term "Observed/Expected Ratio" ("O/E Ratio") refers to the
frequency of CpG dinucleotides within a particular DNA
sequence, and corresponds to the [number of CpG sites / (number
of C bases x number of G bases)] x band length for each
fragment.
The term "CpG island" refers to a contiguous region of genomic
DNA that satisfies the criteria of (1) having a frequency of
CpG dinucleotides corresponding to an "Observed/Expected Ratio"
>0.6, and (2) having a "GC Content" >0.5. CpG islands are
typically, but not always, between about 0.2 to about 1 kb in
length.
The term "methylation state" or "methylation status" refers to
the presence or absence of 5-methylcytosine ("5-mCyt") at one
or a plurality of CpG dinucleotides within a DNA sequence.
Methylation states at one or more particular palindromic CpG
methylation sites (each having two CpG CpG dinucleotide
sequences) within a DNA sequence include "unmethylated,"
"fully-methylated" and "hemi-methylated."
The term "hemi-methylation" or "hemimethylation" refers to the
methylation state of a palindromic CpG methylation site, where
only a single cytosine in one of the two CpG dinucleotide
sequences of the palindromic CpG methylation site is methylated
(e.g., 5'-CC^{M}GG-3' (top strand): 3'-GGCC-5' (bottom strand)).
The term "hypermethylation" refers to the average methylation
state corresponding to an _increased_ presence of 5-mCyt at one
or a plurality of CpG dinucleotides within a DNA sequence of a
test DNA sample, relative to the amount of 5-mCyt found at
corresponding CpG dinucleotides within a normal control DNA
sample.
The term "hypomethylation" refers to the average methylation
state corresponding to a _decreased_ presence of 5-mCyt at one or
a plurality of CpG dinucleotides within a DNA sequence of a

test DNA sample, relative to the amount of 5-mCyt found at corresponding CpG dinucleotides within a normal control DNA sample.

The term "microarray" refers broadly to both "DNA microarrays," and 'DNA chip(s),' as recognized in the art, encompasses all art-recognized solid supports, and encompasses all methods for affixing nucleic acid molecules thereto or synthesis of nucleic acids thereon.

"Genetic parameters" are mutations and polymorphisms of genes and sequences further required for their regulation. To be designated as mutations are, in particular, insertions, deletions, point mutations, inversions and polymorphisms and, particularly preferred, SNPs (single nucleotide polymorphisms). "Epigenetic parameters" are, in particular, cytosine methylations. Further epigenetic parameters include, for example, the acetylation of histones which, however, cannot be directly analyzed using the described method but which, in turn, correlate with the DNA methylation.

The term "bisulfite reagent" refers to a reagent comprising bisulfite, disulfite, hydrogen sulfite or combinations thereof, useful as disclosed herein to distinguish between methylated and unmethylated CpG dinucleotide sequences.

The term "Methylation assay" refers to any assay for determining the methylation state of one or more CpG dinucleotide sequences within a sequence of DNA.

The term "MS.AP-PCR" (Methylation-Sensitive Arbitrarily-Primed Polymerase Chain Reaction) refers to the art-recognized technology that allows for a global scan of the genome using CG-rich primers to focus on the regions most likely to contain CpG dinucleotides, and described by Gonzalgo et al., *Cancer Research* 57:594-599, 1997.

The term "MethyLight™" refers to the art-recognized fluorescence-based real-time PCR technique described by Eads et al., *Cancer Res*. 59:2302-2306, 1999.

The term "HeavyMethyl™" assay, in the embodiment thereof implemented herein, refers to a HeavyMethyl™ refer to the use of methylation specific *blocking* probes covering CpG positions between the amplification primers.

The term "Ms-SNuPE" (Methylation-sensitive Single Nucleotide Primer Extension) refers to the art-recognized assay described by Gonzalgo & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997.

The term "MSP" (Methylation-specific PCR) refers to the art-recognized methylation assay described by Herman et al. *Proc. Natl. Acad. Sci. USA* 93:9821-9826, 1996, and by US Patent No. 5,786,146.

The term "COBRA" (Combined Bisulfite Restriction Analysis) refers to the art-recognized methylation assay described by Xiong & Laird, *Nucleic Acids Res.* 25:2532-2534, 1997.

The term "MCA" (Methylated CpG Island Amplification) refers to the methylation assay described by Toyota et al., *Cancer Res.* 59:2307-12, 1999, and in WO 00/26401A1.

The term "hybridization" is to be understood as a bond of an oligonucleotide to a complementary sequence along the lines of the Watson-Crick base pairings in the sample DNA, forming a duplex structure.

"Stringent hybridization conditions," as defined herein, involve hybridizing at 68°C in 5x SSC/5x Denhardt's solution/1.0% SDS, and washing in 0.2x SSC/0.1% SDS at room temperature, or involve the art-recognized equivalent thereof (e.g., conditions in which a hybridization is carried out at 60°C in 2.5 x SSC buffer, followed by several washing steps at 37°C in a low buffer concentration, and remains stable). Moderately stringent conditions, as defined herein, involve including washing in 3x SSC at 42°C, or the art-recognized equivalent thereof. The parameters of salt concentration and temperature can be varied to achieve the optimal level of identity between the probe and the target nucleic acid. Guidance regarding such conditions is available in the art, for example, by Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, N.Y.; and Ausubel et al. (eds.), 1995, Current Protocols in Molecular Biology, (John Wiley & Sons, N.Y.) at Unit 2.10.

The terms 'sensitivity' and 'specificity' refer to values calculated with reference to a sample set of male patients with an average age of 65 and a mixed ethnic range including caucasian and african american.

Overview:

The present invention provides for molecular genetic markers that have novel utility for the analysis of methylation patterns associated with the development of prostate cell proliferative disorders with a sensitivity of greater than 30% and a specificity of greater than 65%. Said markers may be used for detecting or distinguishing between prostate cell proliferative disorders, thereby providing improved means for the classification and treatment of said disorders. The markers according to the present invention are analysed in the form of a 'panel' wherein the methylation of one or more genetic sequences of the genes

Bisulfite modification of DNA is an art-recognized tool used to assess CpG methylation status. 5-methylcytosine is the most frequent covalent base modification in the DNA of eukaryotic cells. It plays a role, for example, in the regulation of the transcription, in genetic imprinting, and in tumorigenesis. Therefore, the identification of 5-methylcytosine as a component of genetic information is of considerable interest. However, 5-methylcytosine positions cannot be identified by sequencing, because 5-methylcytosine has the same base pairing behavior as cytosine. Moreover, the epigenetic information carried by 5-methylcytosine is completely lost during, *e.g.*, PCR amplification.

The most frequently used method for analyzing DNA for the presence of 5-methylcytosine is based upon the specific reaction of bisulfite with cytosine whereby, upon subsequent alkaline hydrolysis, cytosine is converted to uracil which corresponds to thymine in its base pairing behavior. Significantly, however, 5-methylcytosine remains unmodified under these conditions. Consequently, the original DNA is *converted* in such a manner that methylcytosine, which originally could not be distinguished from cytosine by its hybridization behavior, can now be detected as the only remaining cytosine using standard, art-recognized molecular biological techniques, for example, by amplification and hybridization, or by sequencing. All of these techniques are based on differential base pairing properties, which can now be fully exploited.

The prior art, in terms of sensitivity, is defined by a method comprising enclosing the DNA to be analyzed in an agarose matrix, thereby preventing the diffusion and renaturation of the DNA (bisulfite only reacts with single-stranded DNA), and replacing all precipitation and purification steps with fast dialysis (Olek A, et al., A modified and improved method for bisulfite based cytosine methylation analysis, *Nucleic Acids Res.* 24:5064-6, 1996). It is thus possible to analyze individual cells for methylation status, illustrating the utility and sensitivity of the method. An overview of art-recognized methods for detecting 5-methylcytosine is provided by Rein, T., et al., *Nucleic Acids Res.*, 26:2255, 1998.

The bisulfite technique, barring few exceptions (*e.g.*, Zeschnigk M, et al., *Eur J Hum Genet.* 5:94-98, 1997), is currently only used in research. In all instances, short, specific fragments of a known gene are amplified subsequent to a bisulfite treatment, and either completely sequenced (Olek & Walter, *Nat Genet.* 1997 17:275-6, 1997), subjected to one or more primer extension reactions (Gonzalgo & Jones, *Nucleic Acids Res.*, 25:2529-31, 1997; WO 95/00669; U.S. Patent No. 6,251,594) to analyze individual cytosine positions, or treated by enzymatic digestion (Xiong & Laird, *Nucleic Acids Res.*, 25:2532-4, 1997). Detection by hybridization has also been described in the art (Olek et al., WO 99/28498). Additionally, use of the bisulfite technique for methylation detection with respect to individual genes has been described (Grigg & Clark, *Bioessays*, 16:431-6, 1994; Zeschnigk M, et al., *Hum Mol Genet.*, 6:387-95, 1997; Feil R, et al., *Nucleic Acids Res.*, 22:695-, 1994; Martin V, et al., *Gene*, 157:261-4, 1995; WO 9746705 and WO 9515373).

The present invention provides for the use of the bisulfite technique , in combination with one or more methylation assays, for determination of the methylation status of CpG dinuclotide sequences within sequences from the group consisting of SEQ ID NO: 1 to SEQ ID NO: 4 and SEQ ID NOs: 29 & 30. According to the present invention, determination of the methylation status of CpG dinuclotide sequences within sequences from the group consisting of SEQ ID NO: 1 to SEQ ID NO: 4 and SEQ ID NOs: 29 & 30 has diagnostic and prognostic utility.

Methylation Assay Procedures. Various methylation assay procedures are known in the art, and can be used in conjunction with the present invention. These assays allow for determination of the methylation state of one or a plurality of CpG dinucleotides (*e.g.*, CpG islands) within a DNA sequence. Such assays involve, among other techniques, DNA sequencing of bisulfite-treated DNA, PCR (for sequence-specific amplification), Southern blot analysis, and use of methylation-sensitive restriction enzymes.

For example, genomic sequencing has been simplified for analysis of DNA methylation patterns and 5-methylcytosine distribution by using bisulfite treatment (Frommer et al., *Proc. Natl. Acad. Sci. USA* 89:1827-1831, 1992). Additionally, restriction enzyme digestion of PCR products amplified from bisulfite-converted DNA is used, *e.g.*, the method described by Sadri & Hornsby (*Nucl. Acids Res.* 24:5058-5059, 1996), or COBRA (Combined Bisulfite Restriction Analysis) (Xiong & Laird, *Nucleic Acids Res.* 25:2532-2534, 1997).

COBRA. COBRA analysis is a quantitative methylation assay useful for determining DNA methylation levels at specific gene loci in small amounts of genomic DNA (Xiong & Laird, *Nucleic Acids Res.* 25:2532-2534, 1997). Briefly, restriction enzyme digestion is used to reveal methylation-dependent sequence differences in PCR products of sodium bisulfite-treated DNA. Methylation-dependent sequence differences are first introduced into the genomic DNA by standard bisulfite treatment according to the procedure described by Frommer et al. (*Proc. Natl. Acad. Sci. USA* 89:1827-1831, 1992). PCR amplification of the bisulfite converted DNA is then performed using primers specific for the interested CpG islands, followed by restriction endonuclease digestion, gel electrophoresis, and detection using specific, labeled hybridization probes. Methylation levels in the original DNA sample are represented by the relative amounts of digested and undigested PCR product in a linearly quantitative fashion across a wide spectrum of DNA methylation levels. In addition, this technique can be reliably applied to DNA obtained from microdissected paraffin-embedded tissue samples. Typical reagents (*e.g.*, as might be found in a typical COBRA-based kit) for COBRA analysis may include, but are not limited to: PCR primers for specific gene

(or methylation-altered DNA sequence or CpG island); restriction enzyme and appropriate buffer; gene-hybridization oligo; control hybridization oligo; kinase labeling kit for oligo probe; and radioactive nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery reagents or kits (*e.g.*, precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

Preferably, assays such as "MethyLight•" (a fluorescence-based real-time PCR technique) (Eads et al., *Cancer Res.* 59:2302-2306, 1999), Ms-SNuPE (Methylation-sensitive Single Nucleotide Primer Extension) reactions (Gonzalgo & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997), methylation-specific PCR ("MSP"; Herman et al., *Proc. Natl. Acad. Sci. USA* 93:9821-9826, 1996; US Patent No. 5,786,146), and methylated CpG island amplification ("MCA"; Toyota et al., *Cancer Res.* 59:2307-12, 1999) are used alone or in combination with other of these methods.

MethyLight•. The MethyLight• assay is a high-throughput quantitative methylation assay that utilizes fluorescence-based real-time PCR (TaqMan••) technology that requires no further manipulations after the PCR step (Eads et al., *Cancer Res.* 59:2302-2306, 1999). Briefly, the MethyLight• process begins with a mixed sample of genomic DNA that is converted, in a sodium bisulfite reaction, to a mixed pool of methylation-dependent sequence differences according to standard procedures (the bisulfite process converts unmethylated cytosine residues to uracil). Fluorescence-based PCR is then performed either in an "unbiased" (with primers that do not overlap known CpG methylation sites) PCR reaction, or in a "biased" (with PCR primers that overlap known CpG dinucleotides) reaction. Sequence discrimination can occur either at the level of the amplification process or at the level of the fluorescence detection process, or both.

The MethyLight• assay may be used as a quantitative test for methylation patterns in the genomic DNA sample, wherein sequence discrimination occurs at the level of probe hybridization. In this quantitative version, the PCR reaction provides for unbiased amplification in the presence of a fluorescent probe that overlaps a particular putative

methylation site. An unbiased control for the amount of input DNA is provided by a reaction in which neither the primers, nor the probe overlie any CpG dinucleotides. Alternatively, a qualitative test for genomic methylation is achieved by probing of the biased PCR pool with either control oligonucleotides that do not "cover" known methylation sites (a fluorescence-based version of the "MSP" technique), or with oligonucleotides covering potential methylation sites.

The MethyLight• process can by used with a "TaqMan®" probe in the amplification process. For example, double-stranded genomic DNA is treated with sodium bisulfite and subjected to one of two sets of PCR reactions using TaqMan® probes; e.g., with either biased primers and TaqMan® probe, or unbiased primers and TaqMan® probe. The TaqMan® probe is dual-labeled with fluorescent "reporter" and "quencher" molecules, and is designed to be specific for a relatively high GC content region so that it melts out at about 10°C higher temperature in the PCR cycle than the forward or reverse primers. This allows the TaqMan® probe to remain fully hybridized during the PCR annealing/extension step. As the Taq polymerase enzymatically synthesizes a new strand during PCR, it will eventually reach the annealed TaqMan® probe. The Taq polymerase 5' to 3' endonuclease activity will then displace the TaqMan® probe by digesting it to release the fluorescent reporter molecule for quantitative detection of its now unquenched signal using a real-time fluorescent detection system.

Typical reagents (e.g., as might be found in a typical MethyLight•-based kit) for MethyLight• analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); TaqMan® probes; optimized PCR buffers and deoxynucleotides; and Taq polymerase.

Ms-SNuPE. The Ms-SNuPE technique is a quantitative method for assessing methylation differences at specific CpG sites based on bisulfite treatment of DNA, followed by single-nucleotide primer extension (Gonzalgo & Jones, Nucleic Acids Res. 25:2529-2531, 1997). Briefly, genomic DNA is reacted with sodium bisulfite to convert unmethylated cytosine to uracil while leaving 5-methylcytosine unchanged. Amplification of the desired target sequence is then performed using PCR primers

specific for bisulfite-converted DNA, and the resulting product
is isolated and used as a template for methylation analysis at
the CpG site(s) of interest. Small amounts of DNA can be
analyzed (e.g., microdissected pathology sections), and it
avoids utilization of restriction enzymes for determining the
methylation status at CpG sites.

Typical reagents (e.g., as might be found in a typical Ms-
SNuPE-based kit) for Ms-SNuPE analysis may include, but are not
limited to: PCR primers for specific gene (or methylation-
altered DNA sequence or CpG island); optimized PCR buffers and
deoxynucleotides; gel extraction kit; positive control primers;
Ms-SNuPE primers for specific gene; reaction buffer (for the
Ms-SNuPE reaction); and radioactive nucleotides. Additionally,
bisulfite conversion reagents may include: DNA denaturation
buffer; sulfonation buffer; DNA recovery regents or kit (e.g.,
precipitation, ultrafiltration, affinity column); desulfonation
buffer; and DNA recovery components.

MSP. MSP (methylation-specific PCR) allows for assessing
the methylation status of virtually any group of CpG sites
within a CpG island, independent of the use of methylation-
sensitive restriction enzymes (Herman et al. *Proc. Natl. Acad.
Sci. USA* 93:9821-9826, 1996; US Patent No. 5,786,146).
Briefly, DNA is modified by sodium bisulfite converting all
unmethylated, but not methylated cytosines to uracil, and
subsequently amplified with primers specific for methylated
versus unmethylated DNA. MSP requires only small quantities of
DNA, is sensitive to 0.1% methylated alleles of a given CpG
island locus, and can be performed on DNA extracted from
paraffin-embedded samples. Typical reagents (e.g., as might be
found in a typical MSP-based kit) for MSP analysis may include,
but are not limited to: methylated and unmethylated PCR primers
for specific gene (or methylation-altered DNA sequence or CpG
island), optimized PCR buffers and deoxynucleotides, and
specific probes.

MCA. The MCA technique is a method that can be used to
screen for altered methylation patterns in genomic DNA, and to
isolate specific sequences associated with these changes
(Toyota et al., *Cancer Res.* 59:2307-12, 1999). Briefly,
restriction enzymes with different sensitivities to cytosine
methylation in their recognition sites are used to digest

genomic DNAs from primary tumors, cell lines, and normal tissues prior to arbitrarily primed PCR amplification. Fragments that show differential methylation are cloned and sequenced after resolving the PCR products on high-resolution polyacrylamide gels. The cloned fragments are then used as probes for Southern analysis to confirm differential methylation of these regions. Typical reagents (e.g., as might be found in a typical MCA-based kit) for MCA analysis may include, but are not limited to: PCR primers for arbitrary priming Genomic DNA; PCR buffers and nucleotides, restriction enzymes and appropriate buffers; gene-hybridization oligos or probes; control hybridization oligos or probes.

HeavyMethyl. The HeavyMethyl techniques is a means for selectively amplifying methylated as opposed to non-methylated DNA (or vice versa). Blocker oligonucleotides specific to either methylated or unmenthylated versions of a bisulfite treated target sequence are hybridised to the treated nucleic acids. The sample is then enzymatically amplified, wherein the hybridisation of the blocker oligonucleotides hinders amplification of the nucleic acid strand to which it is bound. Typical reagents (e.g., as might be found in a typical HeavyMethyl-based kit) for HeavyMethyl analysis may include, but are not limited to: methylated or unmethylated blocker oligonucleotides for specific gene (or methylation-altered DNA sequence or CpG island), optimized PCR buffers and deoxynucleotides, and specific probes and primers.

GENOMIC SEQUENCES ACCORDING TO SEQ ID NO: 1 TO SEQ ID NO: 4 AND SEQ ID NOS: 29 & 30, AND TREATED VARIANTS THEREOF ACCORDING TO SEQ ID NO: 5 TO SEQ ID NO: 28, WERE DETERMINED TO HAVE UTILITY FOR DETECTING OR DISTINGUISHING BETWEEN OR AMONG PROSTATE CELL PROLIFERATIVE DISORDERS. .

The present invention is based upon the analysis of methylation levels within one or more genes taken from the group consisting GSTP1, HISTONE H4, PROSTAGLANDIN E2 RECEPTOR, LIM DOMAIN KINASE 1, SEQ ID NO: 29 & ORPHAN NUCLEAR RECEPTOR NR5A2 and their regulatory regions and sequences thereof according to Table 5.

Particular embodiments of the present invention provide a novel application of the analysis of methylation levels and/or patterns within said genes and/or sequences that enables a precise detection, characterisation and/or treatment of prostate cell proliferative disorders. Early detection of prostate cell proliferative disorders is directly linked with disease prognosis, and the disclosed method thereby enables the physician and patient to make better and more informed treatment decisions. The methods disclosed according to the invention enable the detection and characterisation of prostate cell proliferative disorders with improved sensitivity and/or specificity of with a sensitivity of greater than 30% and a specificity of greater than 65%.

FURTHER IMPROVEMENTS

The present invention provides novel uses for genomic sequences selected from the group consisting of SEQ ID NO: 1 TO SEQ ID NO: 4 AND SEQ ID NOS: 29 & 30. Additional embodiments provide modified variants of SEQ ID NO: 1 TO SEQ ID NO: 4 AND SEQ ID NOS: 29 & 30, as well as oligonucleotides and/or PNA-oligomers for analysis of cytosine methylation patterns within SEQ ID NO: 1 TO SEQ ID NO: 4 AND SEQ ID NOS: 29 & 30.

An objective of the invention comprises analysis of the methylation state of one or more CpG dinucleotides within at least one of the genomic sequences selected from the group consisting of SEQ ID NO: 1 TO SEQ ID NO: 4 AND SEQ ID NOS: 29 & 30 and sequences complementary thereto.

The disclosed invention provides treated nucleic acids, derived from genomic SEQ ID NO: 1 to SEQ ID NO: 4 and SEQ ID NOs: 29 & 30 , wherein the treatment is suitable to convert at least one unmethylated cytosine base of the genomic DNA sequence to uracil or another base that is detectably dissimilar to cytosine in terms of hybridization. The genomic sequences in question may comprise one, or more, consecutive or random methylated CpG positions. Said treatment preferably comprises use of a reagent selected from the group consisting of bisulfite, hydrogen sulfite, disulfite, and combinations thereof .In a preferred embodiment of the invention, the objective comprises analysis of a modified nucleic acid comprising a sequence of at least 16 contiguous nucleotide bases in length of a sequence selected from the group

consisting of SEQ ID NO: 5 TO SEQ ID NO: 28, wherein said
sequence comprises at least one CpG, TpA or CpA dinucleotide
and sequences complementary thereto. The sequences of SEQ ID
NO: 5 TO SEQ ID NO: 28 provide modified versions of the nucleic
acid according to SEQ ID NO: 1 TO SEQ ID NO: 4 AND SEQ ID NOS:
29 & 30, wherein the modification of each genomic sequence
results in the synthesis of a nucleic acid having a sequence
that is unique and distinct from said genomic sequence as
follows. For each sense strand genomic DNA, e.g., SEQ ID NO:1,
four converted versions are disclosed. A first version wherein
"C"••"T," but "CpG" remains "CpG" (*i.e.*, corresponds to case
where, for the genomic sequence, all "C" residues of CpG
dinucleotide sequences are methylated and are thus not
converted); a second version discloses the complement of the
disclosed genomic DNA sequence (i.e. *antisense* strand), wherein
"C"••"T," but "CpG" remains "CpG" (*i.e.*, corresponds to case
where, for all "C" residues of CpG dinucleotide sequences are
methylated and are thus not converted). The 'upmethylated'
converted sequences of SEQ ID NO: 1 to SEQ ID NO: 4 and SEQ ID
NOs: 29 & 30 correspond to SEQ ID NO: 5 to SEQ ID NO: 12 and
SEQ ID NO: 21 to 24. A third chemically converted version of
each genomic sequences is provided, wherein "C"••"T" for all
"C" residues, including those of "CpG" dinucleotide sequences
(*i.e.*, corresponds to case where, for the genomic sequences,
all "C" residues of CpG dinucleotide sequences are
unmethylated); a final chemically converted version of each
sequence, discloses the complement of the disclosed genomic DNA
sequence (i.e. *antisense* strand), wherein "C"••"T" for all "C"
residues, including those of "CpG" dinucleotide sequences
(*i.e.*, corresponds to case where, for the complement (*antisense*
strand) of each genomic sequence, all "C" residues of CpG
dinucleotide sequences are unmethylated). The 'downmethylated'
converted sequences of SEQ ID NO: 1 to SEQ ID NO: 4 and SEQ ID
NOs: 29 & 30 correspond to SEQ ID NO: 13 to SEQ ID NO: 20 and
SEQ ID NO: 25 to 28.

In an alternative preferred embodiment, such analysis
comprises the use of an oligonucleotide or oligomer for
detecting the cytosine methylation state within genomic or
pretreated (chemically modified) DNA, according to SEQ ID NO: 1

to SEQ ID NO: 30. Said oligonucleotide or oligomer comprising a nucleic acid sequence having a length of at least nine (9) nucleotides which hybridizes, under moderately stringent or stringent conditions (as defined herein above), to a pretreated nucleic acid sequence according to SEQ ID NO: 5 to SEQ ID NO: 28 and/or sequences complementary thereto, or to a genomic sequence according to SEQ ID NO: 1 to SEQ ID NO: 4 and SEQ ID NOs: 29 & 30 and/or sequences complementary thereto.

Thus, the present invention includes nucleic acid molecules (e.g., oligonucleotides and peptide nucleic acid (PNA) molecules (PNA-oligomers)) that hybridize under moderately stringent and/or stringent hybridization conditions to all or a portion of the sequences SEQ ID NO: 1 to SEQ ID NO: 30, or to the complements thereof. The hybridizing portion of the hybridizing nucleic acids is typically at least 9, 15, 20, 25, 30 or 35 nucleotides in length. However, longer molecules have inventive utility, and are thus within the scope of the present invention.

Preferably, the hybridizing portion of the inventive hybridizing nucleic acids is at least 95%, or at least 98%, or 100% identical to the sequence, or to a portion thereof of SEQ ID NO: 1 to SEQ ID NO: 30, or to the complements thereof.

Hybridizing nucleic acids of the type described herein can be used, for example, as a primer (e.g., a PCR primer), or a diagnostic and/or prognostic probe or primer. Preferably, hybridization of the oligonucleotide probe to a nucleic acid sample is performed under stringent conditions and the probe is 100% identical to the target sequence. Nucleic acid duplex or hybrid stability is expressed as the melting temperature or Tm, which is the temperature at which a probe dissociates from a target DNA. This melting temperature is used to define the required stringency conditions.

For target sequences that are related and substantially identical to the corresponding sequence of SEQ ID NO: 1 to SEQ ID NO: 4 and SEQ ID NOs: 29 & 30 (such as allelic variants and SNPs), rather than identical, it is useful to first establish the lowest temperature at which only homologous hybridization occurs with a particular concentration of salt (e.g., SSC or SSPE). Then, assuming that 1% mismatching results in a 1°C

decrease in the Tm, the temperature of the final wash in the hybridization reaction is reduced accordingly (for example, if sequences having > 95% identity with the probe are sought, the final wash temperature is decreased by 5°C). In practice, the change in Tm can be between 0.5°C and 1.5°C per 1% mismatch.

Examples of inventive oligonucleotides of length X (in nucleotides), as indicated by polynucleotide positions with reference to, *e.g.*, SEQ ID NO:1, include those corresponding to sets (sense and antisense sets) of consecutively overlapping oligonucleotides of length X, where the oligonucleotides within each consecutively overlapping set (corresponding to a given X value) are defined as the finite set of Z oligonucleotides from nucleotide positions:

n to (n + (X-1));

where n=1, 2, 3,…(Y-(X-1));

where Y equals the length (nucleotides or base pairs) of SEQ ID NO:1 (VALUE TO REFLECT LENGTH OF SEQ ID NO:1);

where X equals the common length (in nucleotides) of each oligonucleotide in the set (*e.g.*, X=20 for a set of consecutively overlapping 20-mers); and

where the number (Z) of consecutively overlapping oligomers of length X for a given SEQ ID NO of length Y is equal to Y-(X-1). For example Z= VALUE TO REFLECT LENGTH OF SEQ ID NO:1 -19= VALUE TO REFLECT LENGTH OF SEQ ID NO:1 for either sense or antisense sets of SEQ ID NO:1, where X=20.

Preferably, the set is limited to those oligomers that comprise at least one CpG, TpG or CpA dinucleotide.

Examples of inventive 20-mer oligonucleotides include the following set of VALUE TO REFLECT LENGTH OF SEQ ID NO:1 oligomers (and the antisense set complementary thereto), indicated by polynucleotide positions with reference to SEQ ID NO:1 1-20, 2-21, 3-22, 4-23, 5-24, …… Preferably, the set is limited to those oligomers that comprise at least one CpG, TpG or CpA dinucleotide.

The present invention encompasses, for *each* of SEQ ID NO: 1 to SEQ ID NO: 30 (sense and antisense), multiple consecutively overlapping sets of oligonucleotides or modified

oligonucleotides of length X, where, *e.g.*, X= 9, 10, 17, 20, 22, 23, 25, 27, 30 or 35 nucleotides.

The oligonucleotides or oligomers according to the present invention constitute effective tools useful to ascertain genetic and epigenetic parameters of the genomic sequence corresponding to SEQ ID NO: 1 to SEQ ID NO: 4 and SEQ ID NOs: 29 & 30. Preferred sets of such oligonucleotides or modified oligonucleotides of length X are those consecutively overlapping sets of oligomers corresponding to SEQ ID NO: 1 to SEQ ID NO: 30 (and to the complements thereof). Preferably, said oligomers comprise at least one CpG, TpG or CpA dinucleotide.

Particularly preferred oligonucleotides or oligomers according to the present invention are those in which the cytosine of the CpG dinucleotide (or of the corresponding converted TpG or CpA dinculeotide) sequences is within the middle third of the oligonucleotide; that is, where the oligonucleotide is, for example, 13 bases in length, the CpG, TpG or CpA dinucleotide is positioned within the fifth to ninth nucleotide from the 5'-end.

The oligonucleotides of the invention can also be modified by chemically linking the oligonucleotide to one or more moieties or conjugates to enhance the activity, stability or detection of the oligonucleotide. Such moieties or conjugates include chromophores, fluorophors, lipids such as cholesterol, cholic acid, thioether, aliphatic chains, phospholipids, polyamines, polyethylene glycol (PEG), palmityl moieties, and others as disclosed in, for example, United States Patent Numbers 5,514,758, 5,565,552, 5,567,810, 5,574,142, 5,585,481, 5,587,371, 5,597,696 and 5,958,773. The probes may also exist in the form of a PNA (peptide nucleic acid) which has particularly preferred pairing properties. Thus, the oligonucleotide may include other appended groups such as peptides, and may include hybridization-triggered cleavage agents (Krol et al., *BioTechniques* 6:958-976, 1988) or intercalating agents (Zon, *Pharm. Res.* 5:539-549, 1988). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a chromophore, fluorophor, peptide, hybridization-triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

The oligonucleotide may also comprise at least one art-recognized modified sugar and/or base moiety, or may comprise a modified backbone or non-natural internucleoside linkage.

The oligonucleotides or oligomers according to particular embodiments of the present invention are typically used in 'sets,' which contain at least one oligomer for analysis of each of the CpG dinucleotides of genomic sequence SEQ ID NO: 1 to SEQ ID NO: 4 and SEQ ID NOs: 29 & 30 and sequences complementary thereto, or to the corresponding CpG, TpG or CpA dinucleotide within a sequence of the pretreated nucleic acids according to SEQ ID NO: 5 to SEQ ID NO: 28 and sequences complementary thereto. However, it is anticipated that for economic or other factors it may be preferable to analyze a limited selection of the CpG dinucleotides within said sequences, and the content of the set of oligonucleotides is altered accordingly.

Therefore, in particular embodiments, the present invention provides a set of at least two (2) (oligonucleotides and/or PNA-oligomers) useful for detecting the cytosine methylation state in pretreated genomic DNA (SEQ ID NO: 5 to SEQ ID NO: 28), or in genomic DNA (SEQ ID NO: 1 to SEQ ID NO: 4 and SEQ ID NOs: 29 & 30 and sequences complementary thereto). These probes enable diagnosis, classification and/or therapy of genetic and epigenetic parameters of prostate cell proliferative disorders. The set of oligomers may also be used for detecting single nucleotide polymorphisms (SNPs) in pretreated genomic DNA (SEQ ID NO: 5 to SEQ ID NO: 28), or in genomic DNA (SEQ ID NO: 1 to SEQ ID NO: 4 and SEQ ID NOs: 29 & 30 and sequences complementary thereto).

In preferred embodiments, at least one, and more preferably all members of a set of oligonucleotides is bound to a solid phase.

In further embodiments, the present invention provides a set of at least two (2) oligonucleotides that are used as 'primer' oligonucleotides for amplifying DNA sequences of one of SEQ ID NO: 1 to SEQ ID NO: 30 and sequences complementary thereto, or segments thereof.

It is anticipated that the oligonucleotides may constitute all or part of an "array" or "DNA chip" (i.e., an arrangement of different oligonucleotides and/or PNA-oligomers bound to a

solid phase). Such an array of different oligonucleotide-
and/or PNA-oligomer sequences can be characterized, for
example, in that it is arranged on the solid phase in the form
of a rectangular or hexagonal lattice. The solid-phase surface
may be composed of silicon, glass, polystyrene, aluminum,
steel, iron, copper, nickel, silver, or gold. Nitrocellulose
as well as plastics such as nylon, which can exist in the form
of pellets or also as resin matrices, may also be used. An
overview of the Prior Art in oligomer array manufacturing can
be gathered from a special edition of Nature Genetics (*Nature
Genetics Supplement*, Volume 21, January 1999, and from the
literature cited therein). Fluorescently labeled probes are
often used for the scanning of immobilized DNA arrays. The
simple attachment of Cy3 and Cy5 dyes to the 5'-OH of the
specific probe are particularly suitable for fluorescence
labels. The detection of the fluorescence of the hybridized
probes may be carried out, for example, via a confocal
microscope. Cy3 and Cy5 dyes, besides many others, are
commercially available.

It is particularly preferred that the oligomers according to
the invention are utilised for at least one of: detection
of; detection and differentiation between or among
subclasses of; diagnosis of; prognosis of; treatment of;
monitoring of; and treatment and monitoring of prostate cell
proliferative disorders. This is enabled by use of said sets
for the detection or detection and differentiation of
prostate cell proliferative disorders.

The present invention further provides a method for
ascertaining genetic and/or epigenetic parameters of the genes
GSTP1, HISTONE H4, PROSTAGLANDIN E2 RECEPTOR, LIM DOMAIN
KINASE 1, SEQ ID NO: 29 & ORPHAN NUCLEAR RECEPTOR NR5A2 and
their regulatory regions including genomic sequences according
to SEQ ID NO: 1 to SEQ ID NO: 4 and SEQ ID NOs: 29 & 30
within a subject by analyzing cytosine methylation and single
nucleotide polymorphisms. Said method comprising contacting a
nucleic acid comprising one or more of the genes GSTP1,
HISTONE H4, PROSTAGLANDIN E2 RECEPTOR, LIM DOMAIN KINASE 1,
SEQ ID NO: 29 & ORPHAN NUCLEAR RECEPTOR NR5A2 and their
regulatory regions including genomic sequences according to

SEQ ID NO: 1 to SEQ ID NO: 4 and SEQ ID NOs: 29 & 30 in a biological sample obtained from said subject with at least one reagent or a series of reagents, wherein said reagent or series of reagents, distinguishes between methylated and non-methylated CpG dinucleotides within the target nucleic acid. Preferably, said method comprises the following steps: In the *first step*, a sample of the tissue to be analysed is obtained. The source may be any suitable source, such as cell lines, histological slides, biopsies, tissue embedded in paraffin, bodily fluids, ejaculate, urine, blood and all possible combinations thereof.The DNA is then isolated from the sample. Extraction may be by means that are standard to one skilled in the art, including the use of commercially available kits, detergent lysates, sonification and vortexing with glass beads. Once the nucleic acids have been extracted, the genomic double stranded DNA is used in the analysis.

In the *second step* of the method, the genomic DNA sample is treated in such a manner that cytosine bases which are unmethylated at the 5'-position are converted to uracil, thymine, or another base which is dissimilar to cytosine in terms of hybridization behavior. This will be understood as 'pretreatment' herein.

The above described treatment of genomic DNA is preferably carried out with bisulfite (hydrogen sulfite, disulfite) and subsequent alkaline hydrolysis which results in a conversion of non-methylated cytosine nucleobases to uracil or to another base which is dissimilar to cytosine in terms of base pairing behavior.

In the *third step* of the method, fragments of the pretreated DNA are amplified, using sets of primer oligonucleotides according to the present invention, and an amplification enzyme. The amplification of several DNA segments can be carried out simultaneously in one and the same reaction vessel. Typically, the amplification is carried out using a polymerase chain reaction (PCR). The set of primer oligonucleotides includes at least two oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 16-base-pair long segment of the base sequences of one

or more of SEQ ID NO: 5 to SEQ ID NO: 28 and sequences complementary thereto.

In an alternate embodiment of the method, the methylation status of preselected CpG positions within the nucleic acid sequences comprising one or more of SEQ ID NO: 1 to SEQ ID NO: 4 and SEQ ID NOs: 29 & 30 may be detected by use of methylation-specific primer oligonucleotides. This technique (MSP) has been described in United States Patent No. 6,265,171 to Herman. The use of methylation status specific primers for the amplification of bisulfite treated DNA allows the differentiation between methylated and unmethylated nucleic acids. MSP primers pairs contain at least one primer which hybridizes to a bisulfite treated CpG dinucleotide. Therefore, the sequence of said primers comprises at least one CpG dinucleotide. MSP primers specific for non-methylated DNA contain a "T' at the 3' position of the C position in the CpG. Preferably, therefore, the base sequence of said primers is required to comprise a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence according to one of SEQ ID NO: 5 to SEQ ID NO: 28 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG dinucleotide.

A further preferred embodiment of the method comprises the use of *blocker* oligonucleotides. The use of such blocker oligonucleotides has been described by Yu et al., *BioTechniques* 23:714-720, 1997. Blocking probe oligonucleotides are hybridized to the bisulfite treated nucleic acid concurrently with the PCR primers. PCR amplification of the nucleic acid is terminated at the 5' position of the blocking probe, such that amplification of a nucleic acid is suppressed where the complementary sequence to the blocking probe is present. The probes may be designed to hybridize to the bisulfite treated nucleic acid in a methylation status specific manner. For example, for detection of methylated nucleic acids within a population of unmethylated nucleic acids, suppression of the amplification of nucleic acids which are unmethylated at the position in question would be carried out by the use of blocking probes comprising a 'CpA' or 'TpA' at the position in question, as opposed to a 'CpG' if the suppression of amplification of methylated nucleic acids is desired.

For PCR methods using blocker oligonucleotides, efficient disruption of polymerase-mediated amplification requires that blocker oligonucleotides not be elongated by the polymerase. Preferably, this is achieved through the use of blockers that are 3'-deoxyoligonucleotides, or oligonucleotides derivitized at the 3' position with other than a "free" hydroxyl group. For example, 3'-O-acetyl oligonucleotides are representative of a preferred class of blocker molecule.

Additionally, polymerase-mediated decomposition of the blocker oligonucleotides should be precluded. Preferably, such preclusion comprises either use of a polymerase lacking 5'-3' exonuclease activity, or use of modified blocker oligonucleotides having, for example, thioate bridges at the 5'-terminii thereof that render the blocker molecule nuclease-resistant. Particular applications may not require such 5' modifications of the blocker. For example, if the blocker- and primer-binding sites overlap, thereby precluding binding of the primer (e.g., with excess blocker), degradation of the blocker oligonucleotide will be substantially precluded. This is because the polymerase will not extend the primer toward, and through (in the 5'-3' direction) the blocker—a process that normally results in degradation of the hybridized blocker oligonucleotide.

A particularly preferred blocker/PCR embodiment, for purposes of the present invention and as implemented herein, comprises the use of peptide nucleic acid (PNA) oligomers as blocking oligonucleotides. Such PNA blocker oligomers are ideally suited, because they are neither decomposed nor extended by the polymerase.

Preferably, therefore, the base sequence of said *blocking oligonucleotides* is required to comprise a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence according to one of SEQ ID NO: 5 to SEQ ID NO: 28 and sequences complementary thereto, wherein the base sequence of said oligonucleotides comprises at least one CpG, TpG or CpA dinucleotide.

The fragments obtained by means of the amplification can carry a directly or indirectly detectable label. Preferred are labels in the form of fluorescence labels, radionuclides, or

detachable molecule fragments having a typical mass which can be detected in a mass spectrometer. Where said labels are mass labels, it is preferred that the labeled amplificates have a single positive or negative net charge, allowing for better detectability in the mass spectrometer. The detection may be carried out and visualized by means of, *e.g.*, matrix assisted laser desorption/ionization mass spectrometry (MALDI) or using electron spray mass spectrometry (ESI).

Matrix Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI-TOF) is a very efficient development for the analysis of biomolecules (Karas & Hillenkamp, *Anal Chem.*, 60:2299-301, 1988). An analyte is embedded in a light-absorbing matrix. The matrix is evaporated by a short laser pulse thus transporting the analyte molecule into the vapour phase in an unfragmented manner. The analyte is ionized by collisions with matrix molecules. An applied voltage accelerates the ions into a field-free flight tube. Due to their different masses, the ions are accelerated at different rates. Smaller ions reach the detector sooner than bigger ones. MALDI-TOF spectrometry is well suited to the analysis of peptides and proteins. The analysis of nucleic acids is somewhat more difficult (Gut & Beck, *Current Innovations and Future Trends*, 1:147-57, 1995). The sensitivity with respect to nucleic acid analysis is approximately 100-times less than for peptides, and decreases disproportionally with increasing fragment size. Moreover, for nucleic acids having a multiply negatively charged backbone, the ionization process via the matrix is considerably less efficient. In MALDI-TOF spectrometry, the selection of the matrix plays an eminently important role. For desorption of peptides, several very efficient matrixes have been found which produce a very fine crystallisation. There are now several responsive matrixes for DNA, however, the difference in sensitivity between peptides and nucleic acids has not been reduced. This difference in sensitivity can be reduced, however, by chemically modifying the DNA in such a manner that it becomes more similar to a peptide. For example, phosphorothioate nucleic acids, in which the usual phosphates of the backbone are substituted with thiophosphates, can be converted into a charge-neutral DNA using simple alkylation chemistry (Gut & Beck, *Nucleic Acids*

Res. 23: 1367-73, 1995). The coupling of a charge tag to this modified DNA results in an increase in MALDI-TOF sensitivity to the same level as that found for peptides. A further advantage of charge tagging is the increased stability of the analysis against impurities, which makes the detection of unmodified substrates considerably more difficult.

In the *fourth step* of the method, the amplificates obtained during the third step of the method are analysed in order to ascertain the methylation status of the CpG dinucleotides prior to the treatment.

In embodiments where the amplificates were obtained by means of MSP amplification, the presence or absence of an amplificate is in itself indicative of the methylation state of the CpG positions covered by the primer, according to the base sequences of said primer.

Amplificates obtained by means of both standard and methylation specific PCR may be further analyzed by means of hybridization-based methods such as, but not limited to, array technology and probe based technologies as well as by means of techniques such as sequencing and template directed extension.

In one embodiment of the method, the amplificates synthesised in *step three* are subsequently hybridized to an array or a set of oligonucleotides and/or PNA probes. In this context, the hybridization takes place in the following manner: the set of probes used during the hybridization is preferably composed of at least 2 oligonucleotides or PNA-oligomers; in the process, the amplificates serve as probes which hybridize to oligonucleotides previously bonded to a solid phase; the non-hybridized fragments are subsequently removed; said oligonucleotides contain at least one base sequence having a length of at least 9 nucleotides which is reverse complementary or identical to a segment of the base sequences specified in the present Sequence Listing; and the segment comprises at least one CpG , TpG or CpA dinucleotide.

In a preferred embodiment, said dinucleotide is present in the central third of the oligomer. For example, wherein the oligomer comprises one CpG dinucleotide, said dinucleotide is preferably the fifth to ninth nucleotide from the 5'-end of a 13-mer. One oligonucleotide exists for the analysis of each CpG dinucleotide within the sequence according to SEQ ID NO: 1

to SEQ ID NO: 4 and SEQ ID NOs: 29 & 30, and the equivalent positions within SEQ ID NO: 5 to SEQ ID NO: 28. Said oligonucleotides may also be present in the form of peptide nucleic acids. The non-hybridized amplificates are then removed.The hybridized amplificates are then detected. In this context, it is preferred that labels attached to the amplificates are identifiable at each position of the solid phase at which an oligonucleotide sequence is located.

In yet a further embodiment of the method, the genomic methylation status of the CpG positions may be ascertained by means of oligonucleotide probes that are hybridised to the bisulfite treated DNA concurrently with the PCR amplification primers (wherein said primers may either be methylation specific or standard).

A particularly preferred embodiment of this method is the use of fluorescence-based Real Time Quantitative PCR (Heid et al., *Genome Res.* 6:986-994, 1996; *also see* United States Patent No. 6,331,393) employing a dual-labeled fluorescent oligonucleotide probe (TaqMan™ PCR, using an ABI Prism 7700 Sequence Detection System, Perkin Elmer Applied Biosystems, Foster City, California). The TaqMan™ PCR reaction employs the use of a nonextendible interrogating oligonucleotide, called a TaqMan™ probe, which, in preferred imbodiments, is designed to hybridize to a GpC-rich sequence located between the forward and reverse amplification primers. The TaqMan™ probe further comprises a fluorescent "reporter moiety" and a "quencher moiety" covalently bound to linker moieties (*e.g.*, phosphoramidites) attached to the nucleotides of the TaqMan™ oligonucleotide. For analysis of methylation within nucleic acids subsequent to bisulfite treatment, it is required that the probe be methylation specific, as described in United States Patent No. 6,331,393, (hereby incorporated by reference in its entirety) also known as the MethylLight™ assay. Variations on the TaqMan™ detection methodology that are also suitable for use with the described invention include the use of dual-probe technology (Lightcycler™) or fluorescent amplification primers (Sunrise™ technology). Both these techniques may be adapted in a manner suitable for use with bisulfite treated DNA, and moreover for methylation analysis within CpG dinucleotides.

34

A further suitable method for the use of probe
oligonucleotides for the assessment of methylation by analysis
of bisulfite treated nucleic acids In a further preferred
embodiment of the method, the *fifth step* of the method
comprises the use of template-directed oligonucleotide
extension, such as MS-SNuPE as described by Gonzalgo & Jones,
Nucleic Acids Res. 25:2529-2531, 1997.

In yet a further embodiment of the method, the *fifth step*
of the method comprises sequencing and subsequent sequence
analysis of the amplificate generated in the *third step* of the
method (Sanger F., et al., *Proc Natl Acad Sci USA* 74:5463-5467,
1977).

Best mode

In the most preferred embodiment of the method the
nucleic acids according to SEQ ID NO: 1 to SEQ ID NO: 4 and SEQ
ID NOs: 29 & 30 are isolated and treated according to the
first three steps of the method outlined above, namely:

    a. obtaining, from a subject, a biological sample having
       subject genomic DNA;
    b. extracting or otherwise isolating the genomic DNA;
    c. treating the genomic DNA of b), or a fragment
       thereof, with one or more reagents to convert
       cytosine bases that are unmethylated in the 5-
       position thereof to uracil or to another base that is
       detectably dissimilar to cytosine in terms of
       hybridization properties;

and wherein the subsequent amplification of d) is carried out
in a methylation specific manner, namely by use of methylation
specific primers or *blocking oligonucleotides*, and further
wherein the detection of the amplificates is carried out by
means of a real-time detection probes, as described above.

Wherein the subsequent amplification of d) is carried out
by means of methylation specific primers, as described above,
said methylation specific primers comprise a sequence having a
length of at least 9 nucleotides which hybridizes to a
pretreated nucleic acid sequence according to one of SEQ ID NO:
5 to SEQ ID NO: 28 and sequences complementary thereto, wherein
the base sequence of said oligomers comprises at least one CpG
dinucleotide. Step e) of the method, namely the detection of

the specific amplificates indicative of the methylation status
of one or more CpG positions according to SEQ ID NO: 1 to SEQ
ID NO: 4 and SEQ ID NOs: 29 & 30 is carried out by means of
real-time detection methods as described above.
In an alternative most preferred embodiment of the method the
subsequent amplification of d) is carried out in the presence
of *blocking oligonucleotides*, as described above. Said *blocking
oligonucleotides* comprising a sequence having a length of at
least 9 nucleotides which hybridizes to a pretreated nucleic
acid sequence according to one of SEQ ID NO: 5 to SEQ ID NO: 28
and sequences complementary thereto, wherein the base sequence
of said oligomers comprises at least one CpG, TpG or CpA
dinucleotide. Step e) of the method, namely the detection of
the specific amplificates indicative of the methylation status
of one or more CpG positions according to SEQ ID NO: 1 to SEQ
ID NO: 4 and SEQ ID NOs: 29 & 30 is carried out by means of
real-time detection methods as described above.

14 14 14
Diagnostic and/or Prognostic Assays for prostate cell
proliferative disorders

     The present invention enables diagnosis of events which
are disadvantageous to patients or individuals in which
important genetic and/or epigenetic parameters within one or
more of the genes GSTP1, HISTONE H4, PROSTAGLANDIN E2
RECEPTOR, LIM DOMAIN KINASE 1, SEQ ID NO: 29 & ORPHAN NUCLEAR
RECEPTOR NR5A2 and their regulatory regions including genomic
sequences according to SEQ ID NO: 1 to SEQ ID NO: 4 and SEQ ID
NOs: 29 & 30 may be used as markers. Said parameters obtained
by means of the present invention may be compared to another
set of genetic and/or epigenetic parameters, the differences
serving as the basis for a diagnosis and/or prognosis of events
which are disadvantageous to patients or individuals.
     Specifically, the present invention provides for
diagnostic and/or prognostic cancer assays based on measurement
of differential methylation of one or more CpG dinucleotide
sequences of the genes GSTP1, HISTONE H4, PROSTAGLANDIN E2
RECEPTOR, LIM DOMAIN KINASE 1, SEQ ID NO: 29 & ORPHAN NUCLEAR
RECEPTOR NR5A2 and their regulatory regions including genomic
sequences according to SEQ ID NO: 1 to SEQ ID NO: 4 and SEQ ID

NOs: 29 & 30, or of subregions thereof that comprise such a CpG dinucleotide sequence. Typically, such assays involve obtaining a tissue sample from a test tissue, performing an assay to measure the methylation status of at least one of one or more CpG dinucleotide sequences of the genes GSTP1, HISTONE H4, PROSTAGLANDIN E2 RECEPTOR, LIM DOMAIN KINASE 1, SEQ ID NO: 29 & ORPHAN NUCLEAR RECEPTOR NR5A2 and their regulatory regions including genomic sequences according to SEQ ID NO: 1 to SEQ ID NO: 4 and SEQ ID NOs: 29 & 30 derived from the tissue sample, relative to a control sample, or a known standard and making a diagnosis or prognosis based thereon.

In particular preferred embodiments, inventive oligomers are used to assess the CpG dinucleotide methylation status, such as those based on SEQ ID NO: 1 to SEQ ID NO: 30, or arrays thereof, as well as in kits based thereon and useful for the diagnosis and/or prognosis of prostate cell proliferative disorders.

Kits

Moreover, an additional aspect of the present invention is a kit comprising, for example: a bisulfite-containing reagent; a set of primer oligonucleotides containing at least two oligonucleotides whose sequences in each case correspond, are complementary, or hybridize under stringent or highly stringent conditions to a 16-base long segment of one or more of the genes GSTP1, HISTONE H4, PROSTAGLANDIN E2 RECEPTOR, LIM DOMAIN KINASE 1, SEQ ID NO: 29 & ORPHAN NUCLEAR RECEPTOR NR5A2 and their regulatory regions including genomic and/or treated sequences according to SEQ ID NO: 1 to SEQ ID NO: 30; oligonucleotides and/or PNA-oligomers; as well as instructions for carrying out and evaluating the described method. In a further preferred embodiment, said kit may further comprise standard reagents for performing a CpG position-specific methylation analysis, wherein said analysis comprises one or more of the following techniques: MS-SNuPE, MSP, MethyLight ™, HeavyMethyl™ , COBRA, and nucleic acid sequencing. However, a kit along the lines of the present invention can also contain only part of the aforementioned components.

While the present invention has been described with specificity in accordance with certain of its preferred embodiments, the following example serves only to illustrate the invention and is not intended to limit the invention within the principles and scope of the broadest interpretations and equivalent configurations thereof.

EXAMPLES

The objective of the following study was to analyze the methylation status of prostate cancer markers in different body fluid samples in order to identify the preferred choice of body fluid (urine or serum) for testing and the preferred marker, markers or combinations of markers. The study was run on matched serum and urine sediment samples from 80 patients with an average age of 65 and representative of a number of racial types (caucasian, african american etc..). In each case, genomic DNA was analyzed using the HeavyMethyl or MSP technique after bisulfite conversion.

Urine Sediment was prepared for analysis and bisulphite treated according to the following:

- 200 ul sediment samples were purified using the Magnapure DNA Isolation Kit 1 with a 100 ul elution volume.
- 5 ul HD6 PCR was carried out on the Magnapure Eluate , in order to determine DNA concentration
- 100 ul of the DNA solution was treated using a proprietary bisulfite treatment technique
- 10 ul C3 bisulfite specific quantitative PCR
- 5 ul Merck sulfite test

Serum was prepared for analysis and bisulphite treated according to the following:

- 1 mL serum samples were purified using the Magnapure DNA Large Volume Total nucleic acid with a 100 ul elution volume.
- 5 ul HD6 PCR on Magnapure Eluate – To determine DNA concentration

- 100 ul of the DNA solution was treated using a proprietary bisulfite treatment technique
- 10 ul C3 bisulfite specific quantitative PCR
- 5 ul Merck sulfite test

Single PCR runs were performed on 10 ul of bisulfite treated DNA per sample for each of the markers as described below.

Heavy Methyl Assay of the GSTP1 gene

In the following analysis the methylation status of the gene GSTP1 was analysed by means of methylation specific amplification using the primers according to Table 1 (below).

The sequence of interest is amplified by means of methylation specific primers and a blocker oligonucleotide in order to minimise the unspecific amplification of non methylated DNA. The amplificate is then detected by means of methylation specific Lightcycler probes.

Table 1: Oligonucleotides for MSP – Lightcycler analysis of GSTP1.

SEQ ID NO:	Sequence	Type
31	gggattatttttataaggtt	primer
32	ctctaaaccccatcccc	primer
33	cccatccccaaaaacacaaaccac	blocker
34	CGtCGtCGtAGTtTTCGtt-fluo	probe
35	red640-tAGTGAGTACGCGCGGtt-pho	probe

Reaction conditions:

PCR program

denat at 95°C

   95°C   10min

50 cycles: ramp

  denat at 95°C 10 sec (1°C/s)

annealing 56C°C30 sec (1°C/s) detection

extension 72°C 10 sec (1°C/s)

MSP analysis of the gene HISTONE H4.

In the following analysis the methylation status of the gene HISTONE H4 was analysed by means of methylation specific amplification using the primers according to Table 2 (below).

The sequence of interest is amplified by means of methylation specific primers, the amplificate is then detected by means of methylation specific Taqman probes.

Table 2: Oligonucleotides for MSP - Taqman analysis of HISTONE H4.

SEQ ID NO :	Sequence	Type
36	accgaaaatacgcttcacg	primer
37	gcgttatcgtaaagtattgcgc	primer
38	/56-FAM/cgcgacgaacaaaacgccg/3BHQ_1/	probe

Reaction Conditions:

PCR program

denat at 95°C

95°C   10min

50 cycles: ramp

denat at 95°C   10 sec (20°C/s)

annealing 60C°C45 sec (20°C/s) detection

MSP analysis of the gene PROSTAGLANDIN E2 RECEPTOR.

In the following analysis the methylation status of the gene PROSTAGLANDIN E2 RECEPTOR was analysed by means of methylation

specific amplification using the primers according to Table 3 (below).:

> The sequence of interest is amplified by means of methylation specific primers, the amplificate is then detected by means of methylation specific Taqman probes.

Table 3: Oligonucleotides for MSP - Taqman analysis of PROSTAGLANDIN E2 RECEPTOR

SEQ ID NO :	Sequence	Type
39	cgcgctactccgcataca	primer
40	gaggtaatcgaggcggtcg	primer
41	/56-FAM/cgccaattcatacgccgcacc/3BHQ_1/	probe

PCR program

<u>denat at 95°C</u>

95°C   10min

<u>50 cycles:</u>   ramp

  denat at 95°C   10 sec (20°C/s)

  annealing 60C°C 45 sec (20°C/s) detection

MSP analysis of the gene ORPHAN NUCLEAR RECEPTOR NR5A2.

In the following analysis the methylation status of the gene ORPHAN NUCLEAR RECEPTOR NR5A2 was analysed by means of methylation specific amplification using the primers according to Table 3 (below).

> The sequence of interest is amplified by means of methylation specific primers, the amplificate is then detected by means of methylation specific Taqman probes.

Table 4: Oligonucleotides for MSP - Taqman analysis of ORPHAN
NUCLEAR RECEPTOR NR5A2.

SEQ ID NO :	Sequence	Type
42	ttgtggttcgggaagagac	primer
43	tcccgaactcttcgatcg	primer
44	aactacgcgcaaacccgcga	probe

PCR program
denat at 95°C

  95°C   10min

50 cycles: ramp
  denat at 95°C   10 sec (20°C/s)
  annealing 60C°C 45 sec (20°C/s) detection

Marker Analysis
Results were analyzed qualitatively by scoring amplification as
+/- and quantitatively by determining the percentage of
methylated DNA as a fraction of total DNA calculated using the
C3 bisulfite specific PCR. To measure total methylated DNA, a
100% methylated standard (chemicon SSS1 treated DNA) standard
curve was included in each assay.

Results
For each marker a Receiver Operating Characteristic curve (ROC
curve) of the assay was determined. A ROC is a plot of the true
positive rate against the false positive rate for the different
possible cutpoints of a diagnostic test. It shows the tradeoff
between sensitivity and specificity depending on the selected
cutpoint (any increase in sensitivity will be accompanied by a
decrease in specificity).The area under an ROC curve (AUC) is a
measure for the accuracy of a diagnostic test (the larger the
area the better, optimum is 1, a random test would have a ROC

curve lying on the diagonal with an area of 0.5; for reference: J.P. Egan. Signal Detection Theory and ROC Analysis, Academic Press, New York, 1975).

AUC results:

Serum:

Marker: HevyMethyl GSTP1

AUC: 0.51

Marker: MSP HISTONE H4

AUC: 0.59

Marker: MSP PROSTAGLANDIN E2 RECEPTOR

AUC: 0.52

Marker: MSP ORPHAN NUCLEAR RECEPTOR NR5A2

AUC: 0.50

Urine:

Marker: HeavyMethyl GSTP1

AUC: 0.58

Marker: MSP HISTONE H4

AUC: 0.5

Marker: MSP PROSTAGLANDIN E2 RECEPTOR

AUC: 0.49

Marker: MSP ORPHAN NUCLEAR RECEPTOR NR5A2

AUC: 0.56

In order to provide an accurate detection of prostate cancer it is preferred that a combined analysis of multiple markers is carried out (i.e. a gene panel). For analysis of urine based samples the most preferred combination of markers is GSTP1 , PROSTAGLANDIN E2 RECEPTOR & ORPHAN NUCLEAR RECEPTOR NR5A2 with a sensitivity of 0.37 and a specificity of 0.72.

For analysis of serum based samples the most preferred combination of markers is GSTP1 , HISTONE H4 & ORPHAN NUCLEAR RECEPTOR NR5A2 with a sensitivity of 0.35 and a specificity of 0.75.

MSP analysis of the genes according to Table 1.

In the following analysis the methylation status of a selection of the genes according to Table 5 were analysed by means of MSP TaqMan assays using the primers and probes according to Table 6 (below).

The study was run on 10 samples from prostate carcinoma tissue, 10 from benign prostate hyperplasia tissue and 5 normal prostate tissue samples. Genomic DNA was analyzed using the MSP technique after bisulfite conversion. Total genomic DNA of all samples was bisulfite treated converting unmethylated cytosines to uracil. Methylated cytosines remained conserved. Bisulfite treatment was performed with minor modifications according to the protocol described in Olek et al. (1996).

The sequence of interest was then amplified by means of primers specific for bisulfite treated DNA and the amplificates were detected by means of TaqMan probes using TaqMan and/or Lightcycler platforms. Results are shown in table 7 below.

By combining SEQ ID NO: 29 with SEQ ID NO: 3 a sensitivity of 58% and a specificity of 92% was achieved.

Reagents and cycling conditions were as follows:

SEQ ID NO: 3

Reagent	Final Conc.
Forward Primer	
Reverse Primer	
Probe	
MgCl	
dNTPs	
Water	
Template	
Total	

Temperature (C)	Time (sec)	Rate (C/sec)	Collect	# of Cycles
Denature				1
95	30	120		
Anneal/Ex				50
95	10			
60	45		Single	
Cool				1
40	0.5	20X		

SEQ ID NO: 2

Reagent	MM Volume (ul)
Forward Primer	35.0
Reverse Primer	35.0
Probe	2.8
Water	12.2
Taq	245.0
Total	150.0

Temperature (C)	Time (sec)	# of Cycles
Denature		
91	600	
Annealing		50
95		
60	45	

SEQ ID NO: 30

Reagent	Final Conc.
Forward Primer	0.5 uM
Reverse Primer	
Probe	0.4 uM
MgSO	3.0
Reg Stuff	1.5
Water	
Template	

45

SEQ ID NO: 29

Table 5: Genes & Sequences according to the invention

Gene Name	Genbank Ref. Seq.	Genomic sequence SEQ ID NO:	Treated sequences (methylated) SEQ ID NO:	Treated sequences (unmethylated) SEQ ID NO:
GSTP1	NM_000852	1	5 & 6	13 & 14
HISTONE H4	NM_003495	2	7 & 8	15 & 16

PROSTAGLANDIN E2 RECEPTOR	NM_000958	3	9 & 10	17 & 18
ORPHAN NUCLEAR RECEPTOR NR5A2	NM_003822	4	11 & 12	19 & 20
2308bp non-coding region of 19p13.2*	Contig AC020947.6.1.36838	29	21 & 22	25 & 26
LIM DOMAIN KINASE 1	NM_002314	30	23 & 24	27 & 28

*the sequence is located within the promoter regions of
ENSESTG00002947491 & upstream of upstream of ACP5.

Table 6

Genomic sequence SEQ ID NO:	Forward primer SEQ ID NO:	Reverse primer SEQ ID NO:	Probe SEQ ID NO:
3	45	46	47
2 (assay 4)	48	49	50
2 (assay 1)	51	52	53
30	54	55	56
29	57	58	59

Table 7

Genomic sequence SEQ ID NO:	Sensitivity (%)	Specificity
3	70	88
2 (assay 1)	60	94
30	80	80
29	80	79
2 (assay 4)	60	98

We claim:

1. A method for detecting, or for detecting and distinguishing between or among prostate cell proliferative disorders in a subject with a sensitivity of greater than 30% and a specificity of greater than 65%, said method comprising analysing the methylation pattern of a target nucleic acid comprising one or a combination of sequences taken from the group consisting of SEQ ID Nos: 1-4, 29 & 30 by contacting at least one of said target nucleic acids in a biological sample obtained from said subject with at least one reagent, or series of reagents that distinguishes between methylated and non-methylated CpG dinucleotides.

2. The method of claim 1, wherein prostate carcinoma is distinguished from at least one condition selected from the group consisting of prostate adenoma, normal prostate tissue, non-prostate tissues and non-prostate cell proliferative disorders.

3. A method according to claim 1, comprising:
   -obtaining, from a subject, a biological sample having subject genomic DNA;
   -contacting the genomic DNA, or a fragment thereof, with one reagent or a plurality of reagents for distinguishing between methylated and non methylated CpG dinucleotide sequences within at least one target sequence of the genomic DNA, or fragment thereof, wherein the target sequence comprises, or hybridizes under stringent conditions to, at least 16 contiguous nucleotides of a sequence taken from the group consisting of SEQ ID NO: 1 to SEQ ID NO: 4 and SEQ ID NOs: 29 & 30 , said contiguous nucleotides comprising at least one CpG dinucleotide sequence; and
   -determining, based at least in part on said distinguishing, the methylation state of at least one target CpG dinucleotide sequence, or an average, or a value reflecting an average methylation state of a plurality of target CpG dinucleotide sequences, whereby

detecting, or detecting and distinguishing between or among prostate cell proliferative disorders with a sensitivity of greater than 30% and a specificity of greater than 65% is, at least in part, afforded.

4. The method of claim 3, wherein distinguishing between methylated and non methylated CpG dinucleotide sequences within the target sequence comprises converting unmethylated cytosine bases within the target sequence to uracil or to another base that is detectably dissimilar to cytosine in terms of hybridization properties.

5. The method of claim 3, wherein distinguishing between methylated and non methylated CpG dinucleotide sequences within the target sequence(s) comprises methylation state-dependent conversion or non-conversion of at least one CpG dinucleotide sequence to the corresponding converted or non-converted dinucleotide sequence within a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 28, and contiguous regions thereof corresponding to the target sequence.

6. The method of claim 3, wherein the biological sample is selected from the group consisting of cell lines, histological slides, biopsies, paraffin-embedded tissue, bodily fluids, ejaculate, urine, blood, and combinations thereof.

7. The method of claim 3, wherein distinguishing between methylated and non methylated CpG dinucleotide sequences within the target sequence comprises use of at least one nucleic acid molecule or peptide nucleic acid (PNA) molecule comprising, in each case a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 28, and complements thereof.

8. The method of claim 7, wherein the contiguous sequence comprises at least one CpG, TpG or CpA dinucleotide sequence.

9. The method of claim 7, comprising use of at least two such nucleic acid molecules, or peptide nucleic acid (PNA) molecules.

10. The method of claim 7, comprising use of at least two such nucleic acid molecules, or peptide nucleic acid (PNA) molecules as primer oligonucleotides for the amplification of a sequences selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 28, sequences complementary thereto, and regions thereof that comprise, or hybridize under stringent conditions to the primers.

11. The method of claim 9, comprising use of at least four such nucleic acid molecules, or peptide nucleic acid (PNA) molecules.

12. A method for detecting, or detecting and distinguishing between or among prostate cell proliferative disorders in a subject, comprising:
   a. obtaining, from a subject, a biological sample having subject genomic DNA;
   b. extracting or otherwise isolating the genomic DNA;
   c. treating the genomic DNA of b), or a fragment thereof, with one or more reagents to convert cytosine bases that are unmethylated in the 5-position thereof to uracil or to another base that is detectably dissimilar to cytosine in terms of hybridization properties;
   d. contacting the treated genomic DNA, or the treated fragment thereof, with an amplification enzyme and at least two primers comprising, in each case a contiguous sequence of at least 9 nucleotides that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 28, and complements thereof, wherein the

treated genomic DNA or the fragment thereof is either
amplified to produce at least one amplificate, or is
not amplified; and

e) determining, based on a presence or absence of, or
on a property of said amplificate, the methylation
state of at least one CpG dinucleotide of a sequence
selected from the group consisting SEQ ID NO: 1 to
SEQ ID NO: 4 and SEQ ID NOs: 29 & 30 , or an average,
or a value reflecting an average methylation state of
a plurality of CpG dinucleotides of a sequence
selected from the groups consisting of SEQ ID NO: 1
to SEQ ID NO: 4 and SEQ ID NOs: 29 & 30 , whereby at
least one of detecting, or detecting and
distinguishing between prostate cell proliferative
disorders with a sensitivity of greater than 30% and
a specificity of greater than 65% is at least in
part, afforded.

13.   The method of claim 12, wherein
treating the genomic DNA, or the fragment thereof in c),
comprises use of a reagent selected from the group
consisting of bisulfite, hydrogen sulfite, disulfite, and
combinations thereof.

14.   The method of claim 12, wherein
contacting or amplifying in d) comprises use of at least
one method selected from the group consisting of: use of a
heat-resistant DNA polymerase as the amplification enzyme;
use of a polymerase lacking 5'-3' exonuclease activity;
use of a polymerase chain reaction (PCR); generation of a
amplificate nucleic acid molecule carrying a detectable
labels; and combinations thereof.

15.   The method of claim 14, wherein
the detectable amplificate label is selected from the
label group consisting of: fluorescent labels;
radionuclides or radiolabels; amplificate mass labels
detectable in a mass spectrometer; detachable amplificate
fragment mass labels detectable in a mass spectrometer;
amplificate, and detachable amplificate fragment mass
labels having a single-positive or single-negative net

charge detectable in a mass spectrometer; and combinations thereof.

16.   The method of claim 12, wherein the biological sample obtained from the subject is selected from the group consisting of cell lines, histological slides, biopsies, paraffin-embedded tissue, bodily fluids, ejaculate, urine, blood, and combinations thereof.

17.   The method of claim 12, wherein prostate carcinoma is distinguished from at least one condition selected form the group consisting of prostate adenoma, inflammatory prostate tissue, prostate adenomas with grade 2 dysplasia less than 1 cm, prostate adenomas with grade 3 dysplasia equal to or greater than 1 cm in size, normal prostate tissues, non-prostate normal tissue, body fluids, and non-prostate cancer tissue.

18.   The method of claim 12, further comprising in step d) the use of at least one nucleic acid molecule or peptide nucleic acid molecule comprising in each case a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 28, and complements thereof, wherein said nucleic acid molecule or peptide nucleic acid molecule suppresses amplification of the nucleic acid to which it is hybridized.

19.   The method of claim 18, wherein said nucleic acid molecule or peptide nucleic acid molecule is in each case modified at the 5'-end thereof to preclude degradation by an enzyme having 5'-3' exonuclease activity.

20.   The method of claim 18, wherein said nucleic acid molecule or peptide nucleic acid molecule is in each case lacking a 3' hydroxyl group.

21.   The method of claim 18, wherein the amplification enzyme is a polymerase lacking 5'-3' exonuclease activity.

22.   The method of claim 12, wherein determining in e) comprises hybridization of at least one nucleic acid molecule or peptide nucleic acid molecule in each case comprising a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 28, and complements thereof.

23.   The method of claim 22, wherein at least one such hybridizing nucleic acid molecule or peptide nucleic acid molecule is bound to a solid phase.

24.   The method of claim 22, wherein a plurality of such hybridizing nucleic acid molecules or peptide nucleic acid molecules are bound to a solid phase in the form of a nucleic acid or peptide nucleic acid array selected from the array group consisting of linear or substantially so, hexagonal or substantially so, rectangular or substantially so, and combinations thereof.

25.   The method of claim 22, further comprising extending at least one such hybridized nucleic acid molecule by at least one nucleotide base.

26.   The method of claim 12, wherein determining in e), comprises sequencing of the amplificate.

27.   The method of claim 12, wherein contacting or amplifying in d), comprises use of methylation-specific primers.

28.   The method of claim 12 comprising in d) using primer oligonucleotides comprising one or more CpG; TpG or CpA dinucleotides; and further comprising in e) the use of at least one method selected from the group

consisting of: hybridizing in at least one nucleic acid molecule or peptide nucleid acid molecule comprising a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 28, and complements thereof; hybridizing at least one nucleic acid molecule that is bound to a solid phase and comprises a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 28, and complements thereof; hybridizing at least one nucleic acid molecule comprising a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 28, and complements thereof, and extending at least one such hybridized nucleic acid molecule by at least one nucleotide base; and sequencing in e) of the amplificate.

29.   The method of claim 12 comprising in d) use of at least one nucleic acid molecule or peptide nucleic acid molecule comprising in each case a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 28, and complements thereof, wherein said nucleic acid molecule or peptide nucleic acid molecule suppresses amplification of the nucleic acid to which it is hybridized; and further comprising in e) the use of at least one method selected from the group consisting of: hybridizing in at least one nucleic acid molecule or peptide nucleid acid molecule comprising a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 28, and complements thereof; hybridizing at least

one nucleic acid molecule that is bound to a solid phase and comprises a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 28, and complements thereof; hybridizing at least one nucleic acid molecule comprising a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 28, and complements thereof, and extending at least one such hybridized nucleic acid molecule by at least one nucleotide base; and sequencing in e) of the amplificate.

30.   The method of claim 12, comprising in d) amplification by primer oligonucleotides comprising one or more CpG; TpG or CpA dinucleotides and further comprising in e) hybridizing at least one detectably labeled nucleic acid molecule comprising a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 28.

31.   The method of claim 12, comprising in d) the use of at least one nucleic acid molecule or peptide nucleic acid molecule comprising in each case a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 28, and complements thereof, wherein said nucleic acid molecule or peptide nucleic acid molecule suppresses amplification of the nucleic acid to which it is hybridized, and further comprising in e) hybridizing at least one detectably labeled nucleic acid molecule comprising a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under

moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 28.

32.   A treated nucleic acid derived from genomic SEQ ID NO: 1 to SEQ ID NO: 4 and SEQ ID NOs: 29 & 30 , wherein the treatment is suitable to convert at least one unmethylated cytosine base of the genomic DNA sequence to uracil or another base that is detectably dissimilar to cytosine in terms of hybridization.

33.   A nucleic acid, comprising at least 16 contiguous nucleotides of a treated genomic DNA sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 28, and sequences complementary thereto, wherein the treatment is suitable to convert at least one unmethylated cytosine base of the genomic DNA sequence to uracil or another base that is detectably dissimilar to cytosine in terms of hybridization.

34.   The nucleic acid of claims 32 and 33 wherein the contiguous base sequence comprises at least one CpG, TpG or CpA dinucleotide sequence.

35.   The nucleic acid of claims 32 and 33 wherein the treatment comprises use of a reagent selected from the group consisting of bisulfite, hydrogen sulfite, disulfite, and combinations thereof .

36.   An oligomer, comprising a sequence of at least 9 contiguous nucleotides that is complementary to, or hybridizes under moderately stringent or stringent conditions to a treated genomic DNA sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 28.

37.   The oligomer of Claim 36, comprising at least one CpG , CpA or TpG dinucleotide sequence.

38.   A set of oligomers, comprising at
    least two oligonucleotides according, in each case, to any
    one of Claims 37.

39.   Use of a set of oligomers according, in each case, to
    any one of Claims 36 through 38, as probes for
    determining at least one of a cytosine methylation state,
    or a single nucleotide polymorphism (SNP) of a sequence
    selected from the group consisting of SEQ ID NO: 1 to 4,
    29 & 30 and sequences complementary thereto.

40.   . A kit useful for detecting, or for detecting
    distinguishing between or among prostate cell
    proliferative disorders of a subject, comprising:

-at least one of a bisulfite reagent, or a methylation-
sensitive restriction enzyme; and

-at least one nucleic acid molecule or peptide nucleic acid
molecule comprising, in each case a contiguous sequence at
least 9 nucleotides that is complementary to, or hybridizes
under moderately stringent or stringent conditions to a
sequence selected from the group consisting of SEQ ID 5 to
SEQ ID NO 28, and complements thereof

41.   The kit of claim 40, further
    comprising standard reagents for performing a methylation
    assay selected from the group consisting of MS-SNuPE, MSP,
    MethyLight, HeavyMethyl, COBRA, nucleic acid sequencing,
    and combinations thereof.

42.   The method of any one of claims 1, 12 or 3 comprising
    use of the kit according to claim 41.

43.   Use of a nucleic acid according to claims 32 through
    35, an oligomer according to any one of claims 36 through
    37, a set of oligonucleotides according to claim 38 and a
    kit according to claims 41 and 42 for the detection of,

detection and differentiation between or among subclasses
of prostate cell proliferative disorders.

Abstract

The invention provides methods, nucleic acids and kits for
detecting, or for detecting and distinguishing between or among
prostate cell proliferative disorders. The invention discloses
genomic sequences the methylation patterns of which have
utility for the improved detection of and differentiation
between said class of disorders, thereby enabling the improved
diagnosis and treatment of patients.

<110> Epigenomics AG

<120> METHODS AND NUCLEIC ACIDS FOR THE ANALYSIS OF CpG DINUCLEOTIDE
METHYLATION STATUS ASSOCIATED WITH THE DEVELOPMENT OF PROSTATE
CANCER.

<160> 34

<210> 1
<211> 2501
<212> DNA
<213> Homo Sapiens

<400> 1

```
ttgttgtaca gaatatttca tcacccaggt attatgccga gtacccaata gttctctttt      60
ctgctcctct ccttcctccc atcctgcacc ctggagtcaa ccacagtgtc tgttgtttcc     120
ttgtttgtgt tataagttct catcatttag ctcccactta caagtgagaa catccagtat     180
ttggatttct gttcctgcat tagtttgcta aggataatag cctctagctc catccatgtt     240
cccacaaaag acatgatcta gttctttta atggctgcat taaatgaagt tttaaagata     300
caacataaac accaacctct tccccaccac aaaaatccct tgctgaattt gattacactt     360
aaattaacga gttttgtttc atgaaagact ccttggacaa acttgacagt tgatggaata     420
ggagaagctg tctgtcatgt ctaaagccaa caagagatca atatctagaa taaatggaga     480
tctgcaaatc aacagaaagt aggcagcaaa gccaaagaaa atagcctaag gcacagccac     540
taaaaggaac gtgatcatgt cctttgcagg gacatgggtg gagctggaag ccgttagcct     600
cagcaaactc acacaggaac agaaaaccag cgagaccgca tggtctcact tataagtggg     660
agctgaacaa tgagaacaca tggtcacatg gcggcgatca acacacactg gtgcctgttg     720
agcggggtgc tggggaggga gagtaccagg aagaatagct aagggatact gggcttaata     780
cctgggtgat gggatgatct gtacagcaaa ccatcatggc gcacacacct atgtaacaaa     840
cctgcacatc ctctacatgt accccagaac ttcaaataaa agttggacgg ccaggcgtgg     900
tggctcacgc ctgtaatccc agcactttgg gaagccgagg cgtgcagatc acctaaggtc     960
aggagttcga gaccagcccg gccaacatgg tgaaaccccg tctctactaa aaatacaaaa    1020
atcagccaga tgtggcacgc acctataatt ccacctactc gggaggctga agcagaattg    1080
cttgaacccg agaggcggag gttgcagtga gccgccgaga tcgcgccact gcactccagc    1140
ctgggccaca gcgtgagact acgtcataaa ataaaataaa ataacacaaa ataaaataaa    1200
ataaaataaa ataaaataaa ataaaataaa ataaaataaa ataaaaaaat aaaataaaat    1260
aaaataaaat aaagcaattt cctttcctct aagcggcctc cacccctctc ccctgccctg    1320
tgaagcgggt gtgcaagctc cgggatcgca gcggtcttag ggaatttccc cccgcgatgt    1380
cccggcgcgc cagttcgctg cgcacacttc gctgcggtcc tcttcctgct gtctgtttac    1440
tccctaggcc ccgctgggga cctgggaaag agggaaaggc ttccccggcc agctgcgcgg    1500
cgactccggg gactccaggg cgcccctctg cggccgacgc ccggggtgca gcggccgccg    1560
gggctgggac cggcgggagt ccgcgggacc ctccagaaga gcggccggcg ccgtgactca    1620
gcactggggc ggagcggggc gggaccaccc ttataaggct cggaggccgc gaggccttcg    1680
ctggagtttc gccgccgcag tcttcgccac cagtgagtac gcgcggcccg cgtccccggg    1740
gatggggctc agagctccca gcatggggcc aacccgcagc atcaggcccg ggctcccggc    1800
agggctcctc gcccacctcg agacccggga cgggggccta ggggacccag gacgtcccca    1860
gtgccgttag cggctttcag ggggcccgga gcgcctcggg gagggatggg accccggggg    1920
cggggagggg gggcagactg cgctcaccgc gccttggcat cctcccccgg gctccagcaa    1980
actttctctt gttcgctgca gtgccgccct acaccgtggt ctatttccca gttcgaggta    2040
ggagcatgtg tctggcaggg aagggaggca ggggctgggg ctgcagccca cagccccctcg    2100
cccacccgga gagatccgaa cccccttatc cctccgtcgt gtggcttta ccccgggcct    2160
ccttcctgtt ccccgcctct cccgccatgc ctgctccccg ccccagtgtt gtgtgaaatc    2220
ttcggaggaa cctgtttccc tgttccctcc ctgcactcct gacccctccc cgggttgctg    2280
cgaggcggag tcggcccggt ccccacatct cgtacttctc cctccccgca ggccgctgcg    2340
cggccctgcg catgctgctg gcagatcagg gccagagctg gaaggaggag gtggtgaccg    2400
tggagacgtg gcaggagggc tcactcaaag cctcctgcgt aagtgaccat gcccgggcaa    2460
ggggaggggg tgctgggcct tagggggctg tgactaggat c                        2501
```

<210> 2
<211> 2501
<212> DNA
<213> Homo Sapiens

<400> 2

```
tttgcaaatg gagacatctt cattattcct atagtatcat atgtttttaa agtttgtact      60
cacactttgg gtgataaatg aaggacaaga tccttcccta tccttgtgag gatgactaca     120
gcatgactgg atgggcttgc tatgattttt atctttccct gtgttctcac taccgtttta     180
ttaatctcag ttcttttca cagggtagca cagaatttaa ctagcagaaa gagatccagc     240
catgtagacc agagatttgt ctaagtgacg gcatgtaaga atcaggaagg aaagtttttt     300
gtttaaatac caacaggttc cttccttaaa gcaattatta tttttcaaat ctaacccaca     360
```

```
aggtgatagt atccttaaac caattaaatc agaatctcgg gttggataac ctcaaatatg    420
acttattagc acttcccatt aatcactggt ccttcaggcc tttaagttta cttactagga    480
atctcacttt taataccatc ttatcaactt cagttgtaaa taagagaaca ctcaaaggct    540
gaggaattct cagcggtaaa gctctgccca cgttaagtaa caaaggataa gttagtcttt    600
gttgtgatca ctttgttgta ctgataagct acgtatttct actcaaggat tcaaattctc    660
acctttctca agaattgggc caaaaccgat aaactaaact tatttacggt ccactgatta    720
aaggttgttg cataataagt tcttgctatg ttcagcagtt ggattcacag cgccagaaac    780
ctataactgc ttgactttcc tccccactac actgcgaaaa ttgccccctta aatgtaacta    840
accctaaaac ctcaacagta tcgtggccag gcgtggtggc tcactactgt aataccaaca    900
ttaggcatag gcgaggggat tgaggccagg atatcgaaac tagcctggga aacacacgga    960
gacccggtct ttggaaaaat aattagcctt gcgtggtggt gggcgcgagg ttccggctaa   1020
tcgggaggct acagtgagcc atgatgacac tgcactacag tctgcgcgac ggcccatgtc   1080
agtaagctct ggagcacctg aaacaagttg tgttgggtat tttatttact ggagagcgat   1140
tagtgactga tgcctactta cagcgactag agacgcatgc tccgatagca gcacaaactc   1200
agcaggcgcg aacaaatggt aaagagaaac tgggcaaaca agcatcacgg ctcctcagct   1260
gagaaagtgg gggccctaaa aagggccttt tgttgataga aagggacgct caaccaccga   1320
aaccgtagag ggtgcggccc tggcgcttga gcgcgtagac cacatccatg gcggtgaccg   1380
tcttgcgctt ggcgtgctct gtataggtca cggcgtcccg gatcacgttc tccaggaaca   1440
ccttcagcac cccgcgagtc tcctcgtaga tgaggccgga gatgcgcttc acgccgccgc   1500
ggcgagcaag cgcccggatg gccggcttgg tgatgccctg gatattgtcg cgcagtactt   1560
tacggtggcg cttagcgccg cctttgccaa gacccttccc gcctttgccg cggccagaca   1620
tgacgagcaa gaggagtctc acccaacgct ttgtgaggac tctggcctga ggcagcgcct   1680
ttatacgaca gttggcggac cgaactgaga acctgaaaga agtcggcggg aagtcccgcc   1740
ccggtggggg aggggaaatc taaagggcca aaccgaaata gggggaaaaa aaaagcgagc   1800
ttcttgtttc cgtgttctga attttgtaac gtgcatagta tttgttacc acgttatgag   1860
gctttaaaaa attgcttttg aacgcagaag atatacatca atactgtggg aaatacaaga   1920
aaggacaaga aattaagaaa ctacaatgtt atcccatcac acaggctagt taatcatgta   1980
ttttgcagag cagttgcaca tattttcca agaaaatgta tacagtgttg tatatggagt   2040
tttgtaacct ccttatattg attataattt aaccaatttc tattaaagag ataaaagtga   2100
tgttttggtg tctatgtttc ttaggaatta tcaatagtta taatcagttc cccagcaatt   2160
ttttaatcgg ctgtatttta aaaataatgt tttccacatt caacataaat gtactttttc   2220
tctatacttg ggaccaatat tgaaatttat gattttatta caccaaaatt taaatttat   2280
tacattaata tttaaaattg tattagaggt ctcatgattt ggtactacgg gtctccgcat   2340
tatttccttt ccaaatttcc taatctgttt caccaaggtt tctggacaac tttagagacc   2400
ttttgtgaag tttgaataaa atctcttcga gattttgata attgcattag ctttaggact   2460
taattggaat agaattaaaa tccttaaaac aagctcttat a                       2501
```

<210> 3
<211> 2251
<212> DNA
<213> Homo Sapiens

<400> 3

```
ggaccccgag ccgcccccag gtagccagga gcggcctcag cggcagccgc aaactccagt     60
agccgccgt gctgcccgtg gctggggcgg agggcagcca gagctgggga ccaaggctcc    120
gcgccacctg cgcgcacagc ctcacacctg aacgctgtcc tcccgcagac gagaccggcg    180
ggcactgcaa agctgggact cgtctttgaa ggaaaaaaaa tagcgagtaa gaaatccagc    240
accattcttc actgacccat cccgctgcac ctcttgtttc ccaagttttt gaaagctggc    300
aactctgacc tcggtgtcca aaaatcgaca gccactgaga ccggctttga gaagccgaag    360
atttggcagt ttccagactg agcaggacaa ggtgaaagca ggttggaggc gggtccagga    420
catctgaggg ctgaccctgg gggctcgtga ggctgccacc gctgctgccg ctacaggtga    480
gatggcgttg ggctgacgtt ggggtcaacg ggtagagaac gcagggatgc ggccctcgcc    540
gaagagagcc aagaagggaa gagcgcgctc tccaaattgc ttttgtaact tgttttcagt    600
gagcatttta ttgattcaga atctatcgag aatagcacta gcgagctact tttcccttga    660
gatgggtctt attcatcttg gcaatggagt gagttggatt gtggggagga agaggaatgg    720
gaaaatcagt ttataaatat taatgtcagc aagagtgtgc tgttggcagg acgtatcgcg    780
agcctggaga ttttggtggc cgcagttggt aagtggctac aatccagaaa gtaggatcga    840
gttgctcccc ttgtcttatc agtgtatcgt ttctcgggcg cgggtctaac accttacaag    900
tggtaatttc cgctcacggc agctttgtct ctcttctacc atccccagac ccagccttgc    960
actccaaggc tgcgcaccgc cagccactat catgtccact cccgggggtca attcgtccgc   1020
ctccttgagc cccgaccggc tgaacagccc agtgaccatc ccggcggtga tgttcatctt   1080
cggggtggtg ggcaacctgg tggccatcgt ggtgctgtgc aagtcgcgca aggagcagaa   1140
ggagacgacc ttctacacgc tggtatgtgg gctggctgtc accgacctgt tgggcacttt   1200
gttggtgagc ccggtgacca tcgccacgta catgaagggc caatggcccg ggggccagcc   1260
gctgtgcgag tacagcacct tcattctgct cttcttcagc ctgtccggcc tcagcatcat   1320
ctgcgccatg agtgtcgagc gctacctggc catcaaccat gcctatttct acagccacta   1380
cgtggacaag cgattggcgg cctcacgct ctttgcagtc tatgcgtcca acgtgctctt   1440
ttgcgcgctg cccaacatgg gtctcggtag ctcgcggctg cagtacccag acacctggtg   1500
cttcatcgac tggaccacca acgtgacggc gcacgccgcc tactcctaca tgtacgcggg   1560
cttcagctcc ttcctcattc tcgccaccgt cctctgcaac gtgcttgtgt gcggcgcgct   1620
```

```
gctccgcatg caccgccagt tcatgcgccg cacctcgctg ggcaccgagc agcaccacgc    1680
ggccgcggcc gcctcggttg cctcccgggg ccaccccgct gcctccccag ccttgccgcg    1740
cctcagcgac tttcggcgcc gccggagctt ccgccgcatc gcgggcgccg agatccagat    1800
ggtcatctta ctcattgcca cctccctggt ggtgctcatc tgctccatcc cgctcgtggt    1860
gagtgaccgg ggctgggggcc ctactcggcc tttttctcgc atccacctcc cgcgtccatt    1920
ccccgctccc tgctttccct ctgagtcctt ggcagtgaac gtgtcgcctt taggtcgggg    1980
ctgggattcc cacactgttt ctcagagcag gcccaaccct ctttgaagtc ccaacccta a   2040
cgagatttag caggtgcttt gcccctacat ccccagttt atgttcccgg aagcctgggt      2100
ttctttctcc accgagacag cccttacccc ttgctgcctg acactggccg agtcttccaa     2160
gaaaacccccc cgcccctct gttagacgtg gaggggagcc tgctgtagtg tgacttagcc      2220
cattcctccg tactgtgaac tgtgaactgc a                                    2251
```

<210> 4
<211> 2586
<212> DNA
<213> Homo Sapiens

<400> 4

```
cccgcgcggg cgcgggagta gccccgctgg gcgctcgcag ccgcgggagt caagcccccct    60
ccccaggtgc aggcataaaa gtttatggct cttgaacaat gcggggcaga ggtttttcca    120
agcaacgtct aattggccgc ttctaattaa ggaaagagag gcttccagct ctatggcaac    180
ccaagcaggg cagcttcagg ctaaaggtac tttagaataa taagatcatt ctaagaaatg    240
gaatgtctca ctggacaccc gaacaggttc tctgtcattg gaattggtgt gtactgtact    300
tcaaccagta ctcttgtgtg gagggaggcg acccagtcta ggaaagtcaa ctacagaaag    360
aggtgacctc cgaaaggatt gtcttagcgc tattagaata catgtgacca caccaaaagc    420
ccaggcggac acccgcagcc agctcggatt tggacaattc aacattgctg gcagaactga    480
agggaacaag ttaccccaac cccatcccct gtacgcgtag tgctgagtga gttggggggtg    540
ggaggacagc ggttcgttta ttgccccctt ttaaaatctg agatctgaaa atatggaggt    600
cccattcgtt ttcccagctc ttgattgcca acaaaaaaac aaatcccgct ggctacattt    660
tctcctcatt ccaaaatagc aaccctatgg cttgtattaa gcccttcaga agtttatctc    720
atttgctctg ggccagggag ggaacaatgc taggaaaagt caccggtgct cttccatcct    780
cgcccccttcc agggtgcagg atgtgcgggc cggcgggcct gtgatcccgg aacgcttcct    840
gccatcccct tgcgcgaact tgaaaggact gggaggtgtt gagagcagag ttcagggctg    900
gtgcactctg cggtgctgag tgggcggcgc gcccgggcgc tcaggccggg ggacctgtag    960
tcgccctacc gcggagggga aaatacgtag ctggagggcg tgcgccgtgc gggttgtgat   1020
ccgttacccc atcggtcatc ctggggtctc cccaagcctc taggtagggc tgtgagagtc   1080
ccctagagct gaagccccgg aggctgacct gtgggtctgg ctgctatggg aacccggttg   1140
gtccaaagaa gcctttcttc cgggcacctg gaattccagt ttagtgtggg gcatcgggga   1200
agtggcgctg gggggctggg ttggggacc tcagccggca gctccggaga gggcctaccc   1260
ttggggtcgc tgggtgaggc cggcacgatt cttggctcca aaaggaaagt ttctgcttct   1320
tgttctggcg cgagaagcca aagacttatt ttgagagcgg agagagaaat gttattggta   1380
acgtttttctt tggaaagttc gagaggggtc ttctggacac actacctagt gcccccaaac   1440
cagagaagta gttttttcttt ggtgcctggg ctcagaagtc gccactcact cagcccatgg   1500
ttcgaaatca gcatgggaag cgccgggggca aggcttcgtc ggagactaga ggcctgcctg   1560
tcgggaggag cccctggggg atggggaccc cattctcctg cttgctctgg ttcccacctg   1620
ggacgcctcc gtaggagccc agaaagacga tccactacat ggtcccggga cagagcagcg   1680
cgcccaactt tgagggaact ttgtgcgcct ctctgaggcc ctagctttcc aaggcaccgc   1740
cgtccgttct tctttcccta gaccgaaact ggggaagagt gtgggcgctt ctttgccccg   1800
atgagttcgc ctccccaaac gcctacttcg gctgcaccag agcatctggg aaactctgaa   1860
aggtgcccag gcctcacaca gcagcgtctc cctactcagc ctctgtcttt gggtttttc    1920
aagagagtct ctacctcatg cctcggtctt tcttcgatgt cgggtccccg aggtaggcac   1980
ggagtccctc tgaaagcagt tgcctatctg tgcccctttg gtgtaaagtt agagtttact   2040
ttgttgggggg aaggggaggt agaaaagatc acagttggga aagtgcgctt ttcgccttgt   2100
tcctaaaaca tgcctcaaga ctgtcatcgc gattgttagg agagctatca acgtctaggg   2160
gctataaagg aatttctgaa ccctcggccc ttcccaaacc cccaggttcc taaaacccta   2220
gtgggggtct cttggggctg ggattcaggc tggcaccgct gggaggacct cgcctagcat   2280
cccttttatta atatttcacg aaggcaggct cctgccttct ctggagcctc ttttctcgga   2340
atgttcccaa actctggcta actcactccc ctgtgagcca tcctagggct ctgtggcccg   2400
ggaagagacg cgtcaactcc gcgggtctgc gcgcagtcct tagccgcaaa gtgctgcaag   2460
tgaccccctt gacggcccttt tccgaccgaa gagctcggga accaaagaga aaaaaaataa   2520
ctttattttc aaaagaacaa gtcatcactg cggcgatact gtggcggagg actttggcga   2580
tggggt                                                              2586
```

<210> 5
<211> 2501
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)
```

<400> 5

```
ttgttgtata gaatatttta ttatttaggt attatgtcga gtatttaata gttttttttt      60
ttgttttttt tttttttttt attttgtatt ttggagttaa ttatagtgtt tgttgttttt     120
ttgtttgtgt tataagtttt tattatttag tttttatta taagtgagaa tatttagtat      180
ttggattttt gtttttgtat tagtttgtta aggataatag ttttttagttt tatttatgtt     240
tttataaaag atatgattta gtttttttta atggttgtat taaatgaagt tttaaagata      300
taatataaat attaattttt tttttattat aaaaatttt tgttgaattt gattatattt      360
aaattaacga gttttgtttt atgaaagatt ttttggataa atttgatagt tgatggaata      420
ggagaagttg tttgttatgt ttaaagttaa taagagatta atatttagaa taaatggaga      480
tttgtaaatt aatagaaagt aggtagtaaa gttaaagaaa atagtttaag gtatagttat      540
taaaaggaac gtgattatgt tttttgtagg gatatgggtg gagttggaag tcgttagttt      600
tagtaaattt atataggaat agaaaattag cgagatcgta tggtttatt tataagtggg      660
agttgaataa tgagaatata tggttatatg gcggcgatta atatatattg gtgtttgttg      720
agcgggggtgt tggggaggga gagtattagg aagaatagtt aagggatatt gggtttaata      780
tttgggtgat gggatgattt gtatagtaaa ttattatggc gtatatattt atgtaataaa      840
tttgtatatt ttttatatgt attttagaat tttaaataaa agttggacgg ttaggcgtgg      900
tggtttacgt ttgtaatttt agtattttgg gaagtcgagg cgtgtagatt atttaaggtt      960
aggagttcga gattagttcg gttaatatgg tgaaattcg ttttattaa aaatataaaa     1020
attagttaga tgtggtacgt atttataatt ttatttattc gggaggttga agtagaattg     1080
tttgaattcg agaggcggag gttgtagtga gtcgtcgaga tcgcgttatt gtattttagt     1140
ttgggttata gcgtgagatt acgttataaa ataaaataaa ataatataaa ataaaataaa     1200
ataaaataaa ataaaataaa ataaaataaa ataaaataaa atacaaaaat aaaataaaat     1260
aaaataaaat aaagtaattt ttttttttt aagcggtttt tattttttt ttttgttttg     1320
tgaagcgggt gtgtaagttt cgggatcgta gcggtttag ggaatttttt ttcgcgatgt     1380
ttcggcgcgt tagttcgttg cgtatatttc gttgcggttt ttttttgtt gtttgtttat     1440
ttttaggtt tcgttgggga tttgggaaag agggaaaggt tttttcggtt agttgcgcgg     1500
cgatttcggg gatttttaggg cgtttttttg cggtcgacgt tcggggtgta gcggtcgtcg     1560
gggttggggt cggcgggagt tcgcgggatt ttttagaaga gcggtcggcg tcgtgattta     1620
gtattggggc ggagcggggc gggattattt ttataaggtt cggaggtcgc gaggttttcg     1680
ttggagtttc gtcgtcgtag ttttcgttat tagtgagtac gcgcggttcg cgttttcggg     1740
gatggggttt agagttttta gtatggggtt aattcgtagt attaggttcg ggtttttcggt     1800
agggttttc gtttatttcg agattcggga cgggggttta ggggattag gacgtttta     1860
gtgtcgttag cggtttttag ggggttcgga gcgttcgggg gagggatggg atttcggggg     1920
cggggagggg gggtagattg cgtttatcgc gttttggtat ttttttttcgg gttttagtaa     1980
attttttttt gttcgttgta gtgtcgtttt atatcgtggt ttatttttta gttcgaggta     2040
ggagtatgtg tttggtaggg aagggaggta ggggttgggg ttgtagttta tagttttttcg     2100
tttattcgga gagattcgaa ttttttttatt ttttcgtcgt gtggtttta tttcgggttt     2160
ttttttttgtt tttcgttttt ttcgttatgt ttgtttttcg ttttagtgtt gtgtgaaatt     2220
ttcggaggaa tttgttttt tgttttttttt ttgtattttt gatttttttt cgggttgttg     2280
cgaggcggag tcggttcggt ttttatattt cgtattttt tttttttcgta ggtcgttgcg     2340
cggtttttcg tatgttgttg gtagattagg gttagagttg gaaggaggag gtggtgatcg     2400
tggagacgtg gtaggagggt ttatttaaag ttttttgcgt aagtgattat gttcgggtaa     2460
ggggaggggg tgttgggttt taggggggttg tgattaggat t                        2501
```

<210> 6
<211> 2501
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 6

```
gattttagtt atagttttt aaggtttagt attttttttt tttgttcggg tatggttatt      60
tacgtaggag gttttgagtg agtttttttg ttacgttttt acggttatta tttttttttt     120
ttagtttttgg ttttgatttg ttagtagtat gcgtagggtc gcgtagcggt ttgcggggag     180
ggagaagtac gagatgtggg gatcgggtcg atttcgtttc gtagtaattc gggggaggggt     240
taggagtgta gggagggaat agggaaatag gttttttcga agatttttata taatattggg     300
gcggggagta ggtatggcgg gagaggcggg gaataggaag gaggttcggg gtaaaagtta      360
tacgacggag ggataagggg gttcggattt tttcgggtgg gcgaggggtt gtgggttgta      420
gttttagttt ttgttttttt ttttgttag atatatgttt ttattcgaa ttgggaaata      480
gattacggtg tagggcggta ttgtagcgaa taaagaaaag tttgttggag ttcggggggag     540
gatgttaagg cgcggtgagc gtagttgtt tttttttttc gttttcgggg ttttattttt     600
tttcgaggcg tttcgggttt tttgaaagtc gttaacggta ttgggacgt tttgggtttga     660
ttaggttttc gtttttcggg tt tcgggatgtgg cgaggagttt tgtcgggagt tcgggtttga     720
tgttgcgggt tggttttatg ttgggagttt tgagttttat tttcggggac gcgggtcgcg     780
cgtatttatt ggtggcgaag attgcggcgg cgaaattta gcgaaggttt gcgggttttc      840
gagttttata agggtggttt cgtttcgttt cgttttagtg ttgagttacg gcgtcggtcg      900
```

```
ttttttttgga gggtttcgcg gattttcgtc ggttttagtt tcggcggtcg ttgtatttcg        960
ggcgtcggtc gtagaggggc gttttggagt tttcggagtc gtcgcgtagt tggtcggggga        1020
agttttttttt tttttttag gtttttagcg gggtttaggg agtaaataga tagtaggaag         1080
aggatcgtag cgaagtgtgc gtagcgaatt ggcgcgtcgg gatatcgcgg ggggaaattt         1140
tttaagatcg ttgcgatttc ggagtttgta tattcgtttt ataggtagg ggagaggggt          1200
ggaggtcgtt tagaggaaag gaaattgttt tattttattt tattttattt tattttttta         1260
ttttatttta ttttatttta ttttatttta ttttatttta ttttatttta ttttgtgtta         1320
ttttatttta ttttatgacg tagttttacg ttgtggttta ggttggagtg tagtggcgcg         1380
atttcggcgg tttattgtaa ttttcgtttt tcgggtttaa gtaattttgt tttagttttt         1440
cgagtaggtg gaattatagg tgcgtgttat atttggttga ttttgtatt tttagtagag          1500
acggggtttt attatgttgg tcgggttggt ttcgaatttt tgattttagg tgatttgtac         1560
gtttcggttt tttaaagtgt tgggattata ggcgtgagtt attacgtttg gtcgtttaat         1620
tttttatttga agttttgggg tatatgtaga ggatgtgtag gtttgttata taggtgtgtg        1680
cgttatgatg gtttgttgta tagattattt tattatttag gtattaagtt tagtatttt          1740
tagttatttt ttttggtatt ttttttttttt agtatttcgt ttaataggta ttagtgtgtg        1800
ttgatcgtcg ttatgtgatt atgtgttttt attgtttagt ttttatttat aagtgagatt         1860
atgcggtttc gttggttttt tgttttgtg tgagtttgtt gaggttaacg gttttttagtt         1920
ttatttatgt ttttgtaaag gatatgatta cgtttttttt agtggttgtg ttttaggtta         1980
tttttttttgg ttttgttgtt tattttttgt tgatttgtag attttattt attttagata        2040
ttgattttt gttggtttta gatatgatag atagtttttt ttatttttt aattgttaag          2100
tttgtttaag gagtttttta tgaaataaaa ttcgttaatt taagtgtaat taaatttagt         2160
aagggatttt tgtggtgggg aagaggttgg tgtttatgtt gtatttttaa aattttattt         2220
aatgtagtta ttaaaaagaa ttagattatg tttttgtgg gaatatggat ggagttagag          2280
gttattattt ttagtaaatt aatgtaggaa tagaaattta aatattggat gtttttattt         2340
gtaagtggga gttaaatgat gagaatttat aatataaata aggaaataat agatattgtg         2400
gttgattta gggtgtagga tgggaggaag gagaggagta gaaaagagaa ttattgggta          2460
ttcggtataa tatttgggtg atgaaatatt ttgtataata a                            2501
```

<210> 7
<211> 2501
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 7

```
tttgtaaatg gagatatttt tattattttt atagtattat atgttttaa agtttgtatt          60
tatattttgg gtgataaatg aaggataaga ttttttttta tttttgtgag gatgattata         120
gtatgattgg atgggtttgt tatgattttt attttttttt gtgttttttat tatcgtttta       180
ttaatttttag tttttttta tagggtagta tagaatttaa ttagtagaaa gagatttagt         240
tatgtagatt agagatttgt ttaagtgacg gtatgtaaga attaggaagg aaagttttt          300
gtttaaatat taataggttt ttttttttaaa gtaattatta tttttaaat ttaatttata        360
aggtgatagt attttttaaat taattaaatt agaatttcgg gttggataat tttaaatatg        420
atttattagt attttttatt aattattggt tttttaggtt tttaagttta tttattagga        480
atttttatttt taatattatt ttattaattt tagttgtaaa taagagaata tttaaaggtt        540
gaggaatttt tagcggtaaa gttttgttta cgttaagtaa taaaggataa gttagttttt         600
gttgtgatta ttttgttgta ttgataagtt acgtatttt atttaaggat ttaaattttt          660
atttttttta agaattgggt taaaatcgat aaattaaatt tatttacggt ttattgatta        720
aaggttgttg tataataagt ttttgttatg tttagtagtt ggatttatag cgttagaaat         780
ttataattgt ttgatttttt tttttattat attgcgaaaa ttgtttttta aatgtaatta        840
attttaaaat tttaatagta tcgtggttag gcgtggtggt ttattattgt aatattaata        900
ttaggtatag gcgaggggat tgaggttagg atatcgaaat tagtttggga aatatacgga         960
gattcggttt ttggaaaaat aattagtttt gcgtggtggt gggcgcgagg tttcggttaa        1020
tcgggaggtt atagtgagtt atgatgatat tgtattatag tttgcgcgac ggtttatgtt        1080
agtaagtttt ggagtatttg aaataagttg tgttgggtat tttatttatt ggagagcgat       1140
tagtgattga tgtttatttta tagcgattag agacgtatgt ttcgatagta gtataaattt       1200
agtaggcgcg aataaatggt aaagagaaat tgggtaaata agtattacgg ttttttagtt       1260
gagaaagtgg gggtttttaaa aagggttttt tgttgataga aagggacgtt taattatcga      1320
aatcgtagag ggtgcggttt tggcgtttga gcgcgtagat tatatttatg gcggtgatcg        1380
ttttgcgtttt ggcgtgtttt gtataggtta cggcgtttcg gattacgttt tttaggaata     1440
tttttagtat ttcgcgagtt ttttcgtaga tgaggtcgga gatgcgtttt acgtcgtcgc       1500
ggcgagtaag gcgtcggatg gtcggttggg tgatgttttg gatattgtcg cgtagtattt       1560
tacggtggcg tttagcgtcg tttttgttaa gatttttttc gttttgtcg cggttagata         1620
tgacgagtaa gaggagtttt atttaacgtt ttgtgaggat tttggtttga ggtagcgttt        1680
ttatacgata gttggcggat cgaattgaga atttgaaaga agtcggcggg aagtttcgtt       1740
tcggtggggg agggggaaat taaagggtta aatcgaaata gggggaaaaa aaaagcgagt        1800
ttttttgttt cgtgttttttg attttgtaac gtgtatagta tttgttatt acgttatgag       1860
gttttaaaaa attgttttg aacgtagaag atatatatta atattgtggg aaatataaga       1920
aaggataaga aattaagaaa ttataatgtt atttattat ataggttagt taattatgta         1980
```

ttttgtagag tagttgtata tattttttta agaaaatgta tatagtgttg tatatggagt   2040
tttgtaattt ttttatattg attataattt aattaatttt tattaaagag ataaaagtga   2100
tgttttggtg tttatgtttt ttaggaatta ttaatagtta taattagttt tttagtaatt   2160
ttttaatcgg ttgtatttta aaaataatgt ttttatatt taatataaat gtattttttt   2220
tttatatttg ggattaatat tgaaatttat gattttatta tattaaaatt taaattttat   2280
tatattaata tttaaaattg tattagaggt tttatgattt ggtattacgg gttttcgtat   2340
tatttttttt ttaaattttt taatttgttt tattaaggtt tttggataat tttagagatt   2400
ttttgtgaag tttgaataaa attttttcga gattttgata attgtattag ttttaggatt   2460
taattggaat agaattaaaa tttttaaaat aagtttttat a   2501

<210> 8
<211> 2501
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 8

tataagagtt tgttttaagg attttaattt tattttaatt aagtttttaaa gttaatgtaa   60
ttattaaaat ttcgaagaga ttttatttaa attttataaa aggttttaa agttgtttag   120
aaattttggt gaaatagatt aggaaatttg gaaaggaaat aatgcggaga ttcgtagtat   180
taaattatga gattttaat ataattttaa atattaatgt aataaaattt aaattttggt   240
gtaataaaat tataaatttt aatattggtt ttaagtatag agaaaaagta tattatgtt   300
gaatgtggaa aatattattt ttaaaatata gtcgattaaa aaattgttgg ggaattgatt   360
ataattattg ataatttta agaaatatag atattaaaat attattttta ttttttaat   420
agaaattggt taaattataa ttaatataag gaggttataa aattttatat ataatattgt   480
atatattttt ttggaaaaat atgtgtaatt gttttgtaaa atatatgatt aattagtttg   540
tgtgatggga taatattgta gtttttaat ttttgtttt tttttgtatt ttttatagta   600
ttgatgtata tttttgcgt ttaaaagtaa tttttaaag tttataacg tggtaataaa   660
atattatgta cgttataaaa tttagaatac ggaaataaga agttcgtttt tttttttttt   720
ttatttcggt ttggtttttt agatttttt tttttatcg gggcgggatt tttcgtcgat   780
ttttttagg ttttagttc ggttcgttaa ttgtcgtata aaggcgttgt tttaggttag   840
agttttata aagcgttggg tgagatttt tttgttcgtt atgtttggtc gcggtaaagg   900
cgggaagggt tttggtaaag gcggcgttaa gcgttatcgt aaagtattgc gcgataatat   960
ttagggtatt attaagtcgg ttattcggcg tttgttcgt cgcggcggcg tgaagcgtat   1020
tttcggtttt atttacgagg agattcgcgg ggtgttgaag gtgttttgg agaacgtgat   1080
tcgggacgtc gtgatttata tagagtacgt taagcgtaag acggttatcg ttatggatgt   1140
ggtttacgcg tttaagcgtt agggtcgtat tttttacggt ttcggtggtt gagcgttttt   1200
tttattaat aaaaggtttt ttttagggtt tttatttttt tagttgagga gtcgtgatgt   1260
ttgtttgttt agttttttt tattatttgt tcgcgtttgt tgagtttgtg ttgttatcgg   1320
agtatgcgtt tttagtcgtt gtaagtaggt attagttatt aatcgttttt tagtaaataa   1380
aatatttaat ataatttgtt ttaggtgttt tagagtttat tgatatgggt cgtcgcgtag   1440
attgtagtgt agtgttatta tggtttattg tagtttttcg attagtcgga atttcgcgtt   1500
tattattacg taaggttaat tatttttta aagatcgggt tttcgtgtgt tttttaggtt   1560
agtttcgata ttttggtttt aatttttcg tttatgttta atgttggtat tatagtagtg   1620
agttattacg tttggttacg atattgttga ggttttaggg ttagttatat ttaaggggta   1680
attttcgtag tgtagtgggg aggaaagtta agtagttata ggttttggc gttgtgaatt   1740
taattgttga atatagtaag aatttattat gtaataattt ttaattagtg gatcgtaaat   1800
aagtttagtt tatcggtttt ggtttaattt ttgagaaagg tgagaatttg aatttttgag   1860
tagaaatacg tagtttatta gtataataaa gtgattataa taaagattaa tttatttttt   1920
gttatttaac gtgggtagag ttttatcgtt gagaattttt tagtttttga gtgtttttt   1980
atttataatt gaagttgata agatggtatt aaaagtgaga tttttagtaa gtaaatttaa   2040
aggtttgaag gattagtgat taatgggaag tgttaataag ttatatttga ggttatttaa   2100
ttcgagattt tgatttaatt ggtttaagga tattattatt ttgtgggtta gatttgaaaa   2160
ataataattg ttttaaggaa ggaatttgtt ggtatttaaa taaaaaattt tttttttga   2220
ttttatatg tcgttattta gataaatttt tggtttatat ggttggattt tttttgtta   2280
gttaaatttt gtgttattt gtgaaaaga attgagatta ataaacggt agtgagaata   2340
tagggaaaga taaaaattat agtaagttta tttagttatg ttgtagttat ttttataagg   2400
atagggaagg attttgtttt ttatttatta tttaaagtgt gagtataaat tttaaaaata   2460
tatgatatta taggaataat gaagatgttt ttatttgtaa a   2501

<210> 9
<211> 2251
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 9

```
ggatttcgag tcgttttag gtagttagga gcggtttag cggtagtcgt aaattttagt      60
agtcgttcgt gttgttcgtg gttggggcgg agggtagtta gagttggggga ttaaggtttc    120
gcgttatttg cgcgtatagt tttatatttg aacgttgttt tttcgtagac gagatcggcg    180
ggtattgtaa agttgggatt cgttttgaa ggaaaaaaaa tagcgagtaa gaaatttagt    240
attattttt attgatttat ttcgttgtat ttttgtttt ttaagtttt gaaagttggt    300
aattttgatt tcggtgttta aaaatcgata gttattgaga tcggttttga gaagtcgaag    360
atttggtagt ttttagattg agtaggataa ggtgaaagta ggttggaggc gggtttagga    420
tatttgaggg ttgatttgg gggttcgtga ggttgttatc gttgttgtcg ttataggtga    480
gatggcgttg ggttgacgtt ggggttaacg ggtagagaac gtagggatgc ggttttcgtc    540
gaagagagtt aagaagggaa gagcgcgttt tttaaattgt ttttgtaatt tgttttagt    600
gagtatttta ttgatttaga atttatcgag aatagtatta gcgagttatt ttttttga    660
gatgggtttt atttattttg gtaatggagt gagttggatt gtggggagga agaggaatgg    720
gaaaattagt ttataaatat taatgttagt aagagtgtgt tgttggtagg acgtatcgcg    780
agtttggaga tttggtggt cgtagttggt aagtggttat aatttagaaa gtaggatcga    840
gttgttttt ttgtttttatt agtgtatcgt ttttcgggcg cgggtttaat attttataag    900
tggtaattt cgtttacggt agtttgttt ttttttatt atttttagat ttagttttgt    960
attttaaggt tgcgtatcgt tagttattat tatgtttatt ttcggggtta attcgttcgt   1020
tttttgagt ttcgatcggt tgaatagttt agtgattatt tcggcggtga tgtttatttt   1080
cggggtggtg ggtaatttgg tggttatcgt ggtgttgtgt aagtcgcgta aggagtagaa   1140
ggagacgatt ttttatacgt tggtatgtgg gttggttgtt atcgatttgt tgggtatttt   1200
gttggtgagt tcggtgatta tcgttacgta tatgaagggt taatggttcg ggggttagtc   1260
gttgtgcgag tatagtattt ttattttgtt ttttttagt ttgttcggtt ttagtattat   1320
ttgcgttatg agtgtcgagc gttatttggt tattaattat gtttattttt atagttatta   1380
cgtggataag cgattggcgg gttttacgtt ttttgtagtt tatgcgttta acgtgttttt   1440
ttgcgcgttg tttaatatgg gtttcggtag ttcgcggttg tagtatttag atatttggtg   1500
ttttatcgat tggattatta acgtgacggc gtacgtcgtt tatttttata tgtacgcggg   1560
ttttagtttt ttttttattt tcgttatcgt ttttgtaac gtgtttgtgt gcggcgcgtt   1620
gtttcgtatg tatcgttagt ttatgcgtcg tatttcgttg ggtatcgagt agtattacgc   1680
ggtcgcggtc gtttcggttg ttttcggggg ttatttcgtt gttttttag ttttgtcgcg   1740
ttttagcgat tttcggcgtc gtcggagttt tcgtcgtatc gcgggcgtcg agatttagat   1800
ggttatttta tttattgtta ttttttggt ggtgtttatt tgttttattt cgttcgtggt   1860
gagtgatcgg ggttgggggtt ttattcggtt ttttttcgt atttattttt cgcgtttatt   1920
tttcgttttt tgttttttt ttgagttttt ggtagtgaac gtgtcgtttt taggtcgggg   1980
ttgggatttt tatattgttt tttagagtag gtttaatttt ttttgaagtt ttaattttaa   2040
cgagatttag taggtgtttt gttttatat ttttagttt atgttttcgg aagtttgggt   2100
tttttttttt atcgagatag tttttatttt ttgttgtttg atattggtcg agtttttaa   2160
gaaaattttt cgtttttttt gttagacgtg gaggggagtt tgttgtagtg tgatttagtt   2220
tattttttcg tattgtgaat tgtgaattgt a                                    2251
```

<210> 10
<211> 2251
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 10

```
tgtagtttat agtttatagt acggaggaat gggttaagtt atattatagt aggttttttt      60
ttacgtttaa tagaggggggc gggggggtttt tttggaagat tcggttagtg ttaggtagta    120
aggggtaagg gttgtttcgg tggagaaaga aatttaggtt ttcgggaata taaattgggg    180
gatgtagggg taaagtattt gttaaatttc gttaggggttg ggattttaaa gagggttggg    240
tttgtttga gaaatagtgt gggaattta gtttcgattt aaaggcgata cgtttattgt    300
taaggattta gagggaaagt agggagcggg gaatggacgc gggaggtgga tgcgagaaaa    360
aggtcgagta gggttttagt ttcggttatt tattacgagc gggatggagt agatgagtat    420
tattagggag gtggtaatga gtaagatgat tatttggatt tcggcgttcg cgatgcggcg    480
gaagtttcgg cggcgtcgaa agtcgttgag gcgcggtaag gttggggagg tagcggggtg    540
gtttcgggag gtaatcgagg cggtcgcggt cgcgtggtgt tgttcggtgt ttagcgaggt    600
gcggcgtatg aattggcggt gtatgcggag tagcgcgtcg tatataagta cgttgtagag    660
gacggtggcg agaatgagga aggagttgaa gttcgcgtat atgtaggagt aggcggcgtg    720
cgtcgttacg ttggtggttt agtcgatgaa gtattaggtg tttgggtatt gtagtcgcga    780
gttatcgaga tttatgttgg gtagcgcgta aaagagtacg ttggacgtat agattgtaaa    840
gagcgtgagg ttcgttaatc gtttgtttac gtagtggttg tagaaatagg tatggttgat    900
ggttaggtag cgttcgatat ttatggcgta gatgatgttg aggtcggata ggttgaagaa    960
gagtagaatg aaggtgttgt attcggtatag cggttggttt tcggttatt ggtttttat   1020
gtacgtggcg atgttgttaa taaagtgttt aataggtcgg tgatagttag   1080
tttatatatt agcgtgtaga aggtcgtttt ttttgttttt ttgcgcgatt tgtatagtat   1140
tacgatggtt attaggttgt ttattattttc gaagatgaat attatcgtcg ggatggttat   1200
```

```
tgggttgttt agtcggtcgg ggtttaagga ggcggacgaa ttgatttcgg gagtggatat    1260
gatagtggtt ggcggtgcgt agttttggag tgtaaggttg ggtttgggga tggtagaaga    1320
gagataaagt tgtcgtgagc ggaaattatt atttgtaagg tgttagattc gcgttcgaga    1380
aacgatatat tgataagata aggggagtaa ttcgatttta tttttttggat tgtagttatt    1440
tattaattgc ggttattaaa attttttaggt tcgcgatacg ttttgttaat agtatatttt    1500
tgttgatatt aatatttata aattgatttt tttatttttt ttttttttta taatttaatt    1560
tattttattg ttaagatgaa taagatttat tttaagggaa aagtagttcg ttagtgttat    1620
tttcgataga ttttgaatta ataaaatgtt tattgaaaat aagttataaa agtaatttgg    1680
agagcgcgtt tttttttttt tggttttttt cggcgagggt cgtattttg cgtttttat    1740
tcgttgattt taacgttagt ttaacgttat tttatttgta gcggtagtag cggtggtagt    1800
tttacgagtt tttagggtta gtttttagat gttttggatt cgttttttaat ttgtttttat    1860
tttgtttttgt ttagtttgga aattgttaaa ttttcggttt tttaaagtcg gttttagtgg    1920
ttgtcgatttt-ttggatatcg aggttagagt tgttagtttt taaaaatttg ggaaataaga    1980
ggtgtagcgg gatgggttag tgaagaatgg tgttggatttt tttattcgtt attttttttt    2040
ttttaaagac gagttttagt tttgtagtgt tcgtcggttt cgtttgcggg aggatagcgt    2100
ttaggtgtga ggttgtgcgc gtaggtggcg cggagttttg gtttttagtt ttggttgttt    2160
ttcgttttag ttacgggtag tacgggcggt tattggagtt tgcggttgtc gttgaggtcg    2220
tttttggtta tttggggggcg gttcggggtt t                                  2251
```

```
ttcgcgcggg cgcgggagta gtttcgttgg gcgttcgtag tcgcgggagt taagttttt      60
ttttaggtgt aggtataaaa gtttatggtt tttgaataat gcggggtaga ggttttttta    120
agtaacgttt aattggtcgt ttttaattaa ggaaagagag gtttttagtt ttatggtaat    180
ttaagtaggg tagttttagg ttaaaggtat tttagaataa taagattatt ttaagaaatg    240
gaatgtttta ttggatattc gaataggttt tttgttattg gaattggtgt gtattgtatt    300
ttaattagta ttttgtgtg gagggaggcg atttagttta ggaaagttaa ttatagaaag    360
aggtgattttt cgaaaggatt gttttagcgt tattagaata tatgtgatta tattaaaagt    420
ttaggcggat attcgtagtt agttcggatt tggataattt aatattgttg gtagaattga    480
agggaataag ttattttaat tttatttttt gtacgcgtag tgttgagtga gttggggggtg    540
ggaggatagc ggttcgttta ttgttttttt ttaaaatttg agatttgaaa atatggaggt    600
tttattcgtt tttttagttt ttgattgtta ataaaaaaat aaatttcgtt ggttatattt    660
ttttttatt. ttaaaatagt aattttatgg tttgtattaa gttttttaga agtttatttt    720
atttgtttttg ggttagggag ggaataatgt taggaaaagt tatcggtgtt ttttttatttt    780
cgttttttttt agggtgtagg atgtgcgggt cggcgggttt gtgatttcgg aacgtttttt    840
gttatttttt tgcgcgaatt tgaaaggatt gggaggtgtt gagagtagag tttagggttg    900
gtgtatttttg cggtgttgag tgggcggcgc gttcgggcgt ttaggtcggg ggatttgtag    960
tcgttttatc gcggagggga aaatacgtag ttggagggcg tgcgtcgtgc gggttgtgat   1020
tcgttatttt atcggttatt ttggggtttt tttaagtttt taggtagggt tgtgagagtt   1080
ttttagagtt gaagtttcgg aggttgattt gtgggtttgg ttgttatggg aattcggttg   1140
gtttaaagaa gtttttttttt cgggtatttg gaattttagt ttagtgtggg gtatcgggga   1200
agtggcgttg gggggttggg ttgggggatt ttagtcggta gtttcggaga gggtttattt   1260
ttggggtcgt tgggtgaggt cggtacgatt tttggtttta aaggaaagt ttttgttttt   1320
tgttttggcg cgagaagtta aagatttatt ttgagagcgg agagaaaat gttattggta   1380
acgtttttt tggaaagttc gagagggggtt ttttggatat attatttagt gtttttaaat   1440
tagagaagta gtttttttttt ggtgtttggg tttagaagtc gttatttatt tagtttatgg   1500
ttcgaaatta gtatgggaag cgtcgggggta aggttcgtc ggagattaga ggtttgtttg   1560
tcgggaggag ttttttgggggg atggggattt tatttttttg tttgtttgg ttttttatttg   1620
ggacgttttc gtaggagttt agaaagacga tttattatat ggtttcggga tagagtagcg   1680
cgtttaatttt tgagggaatt ttgtgcgttt tttttgaggtt ttagtttttt aaggtatcgt   1740
cgttcgtttt ttttttttttta gatcgaaatt ggggaagagt gtgggcgttt ttttgtttcg   1800
atgagttcgt tttttttaaac gtttatttcg gttgtattag agtatttggg aaatttttgaa   1860
aggtgtttag gttttatata gtagcgtttt tttatttagt ttttgttttt gggttttttt   1920
aagagagttt ttattttatg tttcggtttt ttttcgatgt cgggttttcg aggtaggtac   1980
ggagttttt tgaaagtagt tgtttatttg tgtttttttg gtgtaaagtt agagtttatt   2040
ttgttggggg aagggggggg agaaaagatt atagtttggga aagtgcgttt ttcgtttttgt   2100
ttttaaaata tgtttttaaga ttgttatcgc gattgttagg agagttatta acgtttaggg   2160
gttataaagg aattttttgaa tttttcggttt tttttttaatt tttaggtttt taaaattttta   2220
gtgggggttt tttgggggttg ggatttaggt tggtatcgtt gggaggattt cgtttagtat   2280
ttttttatta atattttacg aaggtaggtt tttgtttttt ttggagtttt tttttttcgga   2340
atgttttttaa attttggtta atttatttttt ttgtgagtta ttttagggtt ttgtggttcg   2400
ggaagagacg cgttaatttc gcgggtttgc gcgtagtttt tagtcgtaaa gtgttgtaag   2460
tgattttttt gacggttttt ttcgatcgaa gagttcggga attaaagaga aaaaaaataa   2520
```

```
ttttattttt aaaagaataa gttattattg cggcgatatt gtggcggagg attttggcga    2580
tggggt                                                               2586
```

<210> 12
<211> 2586
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 12

```
attttatcgt taaagttttt cgttatagta tcgtcgtagt gatgatttgt tttttttgaaa     60
ataaagttat ttttttttt tttggttttc gagttttcg gtcggaaagg gtcgttaggg     120
gggttatttg tagtattttg cggttaagga ttgcgcgtag attcgcggag ttgacgcgtt     180
tttttcggg ttatagagtt ttaggatggt ttatagggga gtgagttagt tagagtttgg     240
gaatatttcg agaaaagagg tttagagaa ggtaggagtt tgtttcgtg aaatattaat     300
aaagggatgt taggcgaggc tttttttagcg gtgttagttt gaatttttagt tttaagagat     360
ttttattagg gttttaggaa tttggggggtt tgggaagggt cgaggggttta gaaattttt     420
tatagttttt agacgttgat agttttttta ataatcgcga tgatagtttt gaggtatgtt     480
ttaggaataa ggcgaaaagc gtattttttt aattgtgatt ttttttattt ttttttttt     540
taataaagta aattttaatt ttatattaaa ggggtataga taggtaattg ttttttagagg     600
gatttcgtgt ttatttcggg gattcgatat cgaagaaaga tcgaggtatg aggtagagat     660
tttttttgaaa aaatttaaag atagaggttg agtagggaga cgttgttgtg tgaggtttgg     720
gtattttta gagttttta gatgtttttgg tgtagtcgaa gtaggcgtt ggggaggcga     780
atttatcgtg gtaaagaagc gtttatattt ttttttagtt tcggtttagg gaaagaagaa     840
cggacggcgg tgtttggaa agttagggtt ttagagaggc gtataaagtt ttttttaaagt     900
tgggcgcgtt gtttgttttc gggattatgt agtggatcgt ttttttgggt ttttacggag     960
gcgttttagg tgggaattag agtaagtagg agaatggggt ttttatttt tagggggtttt    1020
tttcgatagg taggtttta gttttcgacg aagttttgtt tcggcgtttt ttatgttgat    1080
ttcgaattat gggttgagtg agtggcgatt tttgagttta ggtattaaag aaaaattatt    1140
tttttggttt gggggtatta ggtagtgtgt ttagaagatt tttttcgaat ttttttaaaga    1200
aaacgttatt aataatattt ttttttttcgt ttttaaaata agttttttggt ttttcgcgtt    1260
agaataagaa gtagaaattt ttttttttgga gttaagaatc gtgtcggttt tatttagcga    1320
ttttaagggt aggttttttt cggagttgtc ggttgaggtt ttttaattta gtttttttagc    1380
gttattttttt cgatgtttta tattaaattg gaatttttagg tgttcggaag aaaggttttt    1440
ttggattaat cgggtttttta tagtagttag atttataggt tagtttttcgg ggtttttagtt    1500
ttaggggatt tttatagttt tatttagagg tttggggggaa ttttaggatg atcgatgggg    1560
taacggatta taattcgtac ggcgtacgtt ttttagttac gtatttttt tttcgcggta    1620
gggcgattat aggttttttcg gtttgagcgt tcgggcgcgt cgtttatttta gtatcgtaga    1680
gtgtattagt tttgaatttt gttttttaata ttttttagtt tttttaagtt cgcgtaaggg    1740
gatggtagga agcgtttcgg gattataggt tcgtcggttc gtatatttttg tatttttggaa    1800
ggggcgagga tggaagagta tcggtgattt ttttttagtat tgtttttttt ttggtttaga    1860
gtaaatgaga taaatttttg aagggtttaa tataagttat agggttgtta ttttggaatg    1920
aggagaaaat gtagttagcg ggatttgttt ttttgttggt aattaagagt tgggaaaacg    1980
aatgggattt ttatatttt agatttttaga ttttaaaagg gggtaataaa cgaatcgttg    2040
ttttttttatt tttaatttat ttagtattac gcgtataggg gatggggttg gggtaatttg    2100
tttttttttag ttttgttagt aatgttgaat tgtttaaatt cgagttggtt gcgggtgttc    2160
gtttgggttt ttggtgtggt tatatgtatt ttaatagcgt taagataatt ttttcggagg    2220
ttatttttttt ttgtagttga ttttttttaga ttgggtcgtt ttttttttata taagagtatt    2280
ggttgaagta tagtatatat taatttttaat gatagagaat ttgttcgggt gtttagtgag    2340
atattttatt ttttagaatg attttattat tttaaagtat ttttagtttg aagttgtttt    2400
gtttgggttg ttatagagtt ggaagttttt tttttttttaa ttagaagcgg ttaattagac    2460
gttgtttgga aaaattttttg tttcgtattg tttaagagtt ataaattttt atgtttgtat    2520
ttggggaggg ggtttgattt tcgcggttgc gagcgtttag cggggttatt ttcgcgttcg    2580
cgcggg                                                               2586
```

<210> 13
<211> 2501
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 13

```
ttgttgtata gaatattta ttatttaggt attatgttga gtatttaata gttttttttt     60
ttgtttttttt ttttttttt attttgtatt ttggagttaa ttatagtgtt tgttgttttt    120
ttgtttgtgt tataagtttt tattatttag ttttattta taagtgagaa tatttagtat    180
```

```
ttggatttt  gttttgtat  tagtttgtta  aggataaatag  ttttagttt  tattatgtt       240
tttataaaag  atatgattta  gtttttttta  atggttgtat  taaatgaagt  tttaaagata     300
taatataaat  attaatttt  ttttattat  aaaaatttt  tgttgaattt  gattatattt        360
aaattaatga  gtttgtttt  atgaaagatt  ttttggataa  atttgatagt  tgatggaata      420
ggagaagttg  tttgttatgt  ttaaagttaa  taagagatta  atatttagaa  taaatggaga     480
tttgtaaatt  aatagaaagt  aggtagtaaa  gttaaagaaa  atagtttaag  gtatagttat     540
taaaaggaat  gtgattatgt  tttttgtagg  gatatgggtg  gagttggaag  ttgttagttt     600
tagtaaattt  atataggaat  agaaaattag  tgagattgta  tggttttatt  tataagtggg     660
agttgaataa  tgagaatata  tggttatatg  gtggtgatta  atatatattg  gtgtttgttg     720
agtgggggtgt  tggggagggga  gagtattagg  aagaatagtt  aagggatatt  gggtttaata   780
tttgggtgat  gggatgattt  gtatagtaaa  ttattatggt  gtatatattt  atgtaataaa     840
tttgtatatt  ttttatatgt  attttagaat  tttaaataaa  agttggatgg  ttaggtgtgg     900
tggtttatgt  ttgtaatttt  agtattttgg  gaagttgagg  tgtgtagatt  atttaaggtt     960
aggagtttga  gattagtttg  gttaatatgg  tgaaattttg  tttttattaa  aaatataaaa    1020
attagttaga  tgtggtatgt  atttataatt  ttatttattt  gggaggttga  agtagaattg    1080
tttgaatttg  agaggtggag  gttgtagtga  gttgttgaga  ttgtgttatt  gtattttagt    1140
ttgggttata  gtgtgagatt  atgttataaa  ataaaataaa  ataatataaa  ataaaataaa    1200
ataaaataaa  ataaaataaa  ataaaataaa ·ataaaataaa  ataaaaaaat  aaaataaaat    1260
aaaataaaat  aaagtaattt  ttttttttt  aagtggtttt  tattttttt  ttttgttttg      1320
tgaagtgggt  gtgtaagttt  tgggattgta  gtggtttag  ggaatttttt  tttgtgatgt      1380
tttggtgtgt  tagtttgttg  tgtatatttt  gttgtggttt  ttttttgtt  gtttgtttat      1440
tttttaggtt  ttgttgggga  tttgggaaag  agggaaaggt  tttttggtt  agttgtgtgg      1500
tgattttggg  gattttaggg  tgtttttttg  tggttgatgt  ttggggtgta  gtggttgttg     1560
gggttggggt  tggtgggagt  ttgtgggatt  ttttagaaga  gtggttggtg  ttgtgattta     1620
gtattgggt  ggagtggggt  gggattattt  ttataaggt  tggaggttgt  gaggtttttg       1680
ttggagtttt  gttgttgtag  ttttgttat  tagtgagtat  gtgtggtttg  tgttttggg       1740
gatggggttt  agagttttta  gtatggggtt  aatttgtagt  attaggtttg  ggtttttggt     1800
agggtttttt  gtttattttg  agatttggga  tggggggttta  ggggatttag  gatgtttta     1860
gtgttgttag  tggttttag  ggggtttgga  gtgtttgggg  gagggatggg  atttgggggg      1920
tggggagggg  gggtagattg  tgtttattgt  gttttggtat  ttttttttgg  gtttagtaa     1980
atttttttt  gtttgttgta  gtgttgtttt  atattgtggt  ttattttta  gtttgaggta     2040
ggagtatgtg  tttggtaggg  aagggaggta  ggggttgggg  ttgtagttta  tagttttttg    2100
tttatttgga  gagatttgaa  ttttttttatt  tttttgttgt  gtggttttta  ttttgggttt   2160
ttttttgtt  tttgtttttt  tttgttatgt  ttgttttttg  tttagtgtt  gtgtgaaatt      2220
tttggaggaa  tttgtttttt  tgtttttttt  ttgtattttt  gatttttttt  tgggttgttg    2280
tgaggtggag  ttggtttggt  ttttatattt  tgtattttt  ttttttgta  ggttgttgtg      2340
tggtttgtg  tatgttgttg  gtagattagg  gttagagttg  gaaggaggag  gtggtgattg      2400
tggagatgtg  gtaggagggt  ttatttaaag  tttttgtgt  aagtgattat  gtttgggtaa     2460
ggggagggggg  tgttgggttt  taggggggttg  tgattaggat  t                       2501
```

<210> 14
<211> 2501
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 14

```
gattttagtt  atagtttttt  aaggtttagt  atttttttt  tttgtttggg  tatggttatt      60
tatgtaggag  gttttgagtg  agtttttttg  ttatgttttt  atggttatta  ttttttttt     120
ttagtttttgg  ttttgatttg  ttagtagtat  gtgtagggtt  gtgtagtggt  ttgtggggag    180
ggagaagtat  gagatgtggg  gattgggttg  attttgtttt  gtagtaattt  ggggagggggt   240
taggagtgta  gggagggaat  agggaaatag  gtttttttga  agatttata  taatattggg     300
gtggggagta  ggtatggtgg  gagaggtggg  gaataggaag  gaggtttggg  gtaaaagtta     360
tatgatggag  ggataagggg  gtttggattt  ttttgggtgg  gtgaggggtt  gtgggttgta    420
gttttagttt  ttgtttttt  ttttgttag  atatatgttt  ttattttgaa ·ttgggaaata      480
gattatggtg  tagggtggta  ttgtagtgaa  taaagaaaag  tttgttggag  tttggggggag    540
gatgttaagg  tgtggtgagt  gtagtttgtt  tttttttttt  gttttgggg  ttttattttt      600
ttttgaggtg  ttttgggttt  tttgaaagtt  gttaatggta  ttgggatgt  tttgggtttt      660
ttaggttttt  gttttgggtt  ttgaggtggg  tgaggagttt  tgttgggagt  ttgggtttga     720
tgttgtgggt  tggttttatg  ttgggagttt  tgagttttat  tttttgggat  gtgggttgtg     780
tgtatttatt  ggtggtgaag  attgtggtgg  tgaaatttta  gtgaaggttt  tgtggttttt     840
gagtttttata  agggtggttt  tgttttgttt  tgttttagtg  ttggtggttg  gtgttggttg    900
ttttttttgga  gggttttgtg  gattttttgtt  ggttttagtt  ttggtggttg  ttgtattttg   960
ggtgttggtt  gtagagggggt  gttttggagt  ttttggagtt  gttgtgtagt  tggttggggga  1020
agttttttttt  ttttttttag  gttttagtg  gggtttaggg  agtaaataga  tagtaggaag    1080
aggattgtag  tgaagtgtgt  gtagtgaatt  ggtgtgttgg  gatattgtgg  ggggaaattt    1140
tttaagattg  ttgtgatttt  ggagtttgta  tatttgtttt  atagggtagg  ggagagggggt   1200
ggaggttgtt  tagaggaaag  gaaattgttt  tattttattt  tattttattt  tatttttta    1260
```

```
ttttatttta ttttatttta ttttatttta ttttatttta ttttatttta ttttgtgtta   1320
ttttatttta ttttatgatg tagtttttatg ttgtggttta ggttggagtg tagtggtgtg   1380
attttggtgg tttattgtaa ttttttgtttt ttgggtttaa gtaattttgt tttagttttt   1440
tgagtaggtg gaattatagg tgtgtgttat atttggttga ttttttgtatt tttagtagag   1500
atggggtttt attatgttgg ttgggttggt tttgaatttt tgattttagg tgatttgtat   1560
gttttggttt tttaaagtgt tgggattata ggtgtgagtt attatgtttg gttgtttaat   1620
ttttatttga agttttgggg tatatgtaga ggatgtgtag gtttgttata .taggtgtgtg   1680
tgttatgatg gtttgttgta tagattattt tattatttag gtattaagtt tagtattttt   1740
tagttatttt ttttggtatt ttttttttttt agtattttgt ttaataggta ttagtgtgtg   1800
ttgattgttg ttatgtgatt atgtgttttt attgtttagt ttttatttat aagtgagatt   1860
atgtggtttt gttggttttt tgttttgtgtg tgagtttgtt gaggttaatg gtttttagtt   1920
ttatttatgt ttttgtaaag gatatgatta tgtttttttt agtggttgtg tttttaggtta   1980
ttttttttgg ttttgttgtt tattttttgt tgatttgtag atttttattt attttagata   2040
ttgattttttt gttggtttta gatatgatag atagtttttt ttatttattt aattgttaag   2100
tttgtttaag gagttttttta tgaaataaaa tttgttaatt taagtgtaat taaatttagt   2160
aagggatttt tgtggtgggg aagaggttgg tgtttatgtt gtatttttaa aattttatttt   2220
aatgtagtta ttaaaaagaa ttagattatg ttttttgtgg gaatatggat ggagttagag   2280
gttattattt ttagtaaatt aatgtaggaa tagaaattta aatattggat gtttttattt   2340
gtaagtggga gttaaatgat gagaatttat aatataaata aggaaataat agatattgtg   2400
gttgatttta gggtgtagga tgggaggaag gagaggagta gaaaagagaa ttattgggta   2460
tttggtataa tatttgggtg atgaaatatt ttgtataata a                       2501
```

<210> 15
<211> 2501
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 15

```
tttgtaaatg gagatatttt tattattttt atagtattat atgtttttaa agtttgtatt     60
tatatttggg gtgataaatg aaggataaga ttttttttta tttttgtgag gatgattata    120
gtatgattgg atgggtttgt tatgattttt attttttttt gtgttttttat tattgtttta    180
ttaattttag tttttttttta tagggtagta tagaatttaa ttagtagaaa gagatttagt    240
tatgtagatt agagatttgt ttaagtgatg gtatgtaaga attaggaagg aaagttttttt    300
gtttaaatat taataggttt ttttttttaaa gtaattatta ttttttaaat ttaatttata    360
aggtgatagt atttttaaat taattaaatt agaattttgg gttggataat tttaaatatg    420
atttattagt attttttatt aattattggt ttttttaggtt tttaagttta tttattagga    480
atttttatttt taatattatt ttattaattt tagttgtaaa taagagaata tttaaaggtt    540
gaggaatttt tagtggtaaa gtttttgttta tgttaagtaa taaaggataa gttagttttt    600
gttgtgatta ttttgttgta ttgataagtt atgtattttt atttaaggat ttaaattttt    660
atttttttta agaattgggt taaaattgat aaattaaatt tatttatggt ttattgatta    720
aaggttgttg tataataagt ttttgttatg tttagtagtt ggatttatag tgttagaaat    780
ttataattgt ttgatttttt ttttttattat attgtgaaaa ttgttttttta aatgtaatta    840
attttaaaat tttaatagta ttgtggttag gtgtggtggt ttattattgt aatattaata    900
ttaggtatag gtgagggat tgaggttagg atattgaaat tagtttggga aatatatgga     960
gatttggttt ttggaaaaat aattagtttt gtgtggtggt gggtgtgagg ttttggttaa   1020
ttgggaggtt atagtgagtt atgatgatat tgtattatag tttgtgtgat ggtttatgtt   1080
agtaagtttt ggagtatttg aaataagttg tgttgggtat tttatttatt ggagagtgat   1140
tagtgattga tgtttatta tagtgattag agatgtatgt tttgatagta gtataaattt   1200
agtaggtgtg aataaatggt aaagagaaat tgggtaaata agtattatgg ttttttagtt   1260
gagaaagtgg gggtttttaaa aagggttttt tgttgataga aagggatgtt taattattga   1320
aattgtagag ggtgtggttt tggtgtttga gtgtgtagat tatatttatg gtggtgattg   1380
ttttgtgttt ggtgtgtttt gtataggtta tggtgttttg gattatgttt tttaggaata   1440
ttttttagtat tttgtgagtt tttttgtaga tgaggttgga gatgtgtttt atgttgttgt   1500
ggtgagtaag gtgttggatg gttggtttgg tgatgttttg gatattgttg tgtagtattt   1560
tatggtggtg tttagtgttg tttttgttaa gattttttttt gttttgttg tggttagata   1620
tgatgagtaa gaggagtttt atttaatgtt ttgtgaggat tttggtttga ggtagtgttt   1680
ttatatgata gttggtggat tgaattgaga atttgaaaga agttggtggg aagtttttgtt   1740
ttggtggggg aggggaaatt taaagggtta aattgaaata ggggaaaaa aaaagtgagt   1800
ttttttgtttt tgtgttttga atttttgtaat gtgtatagta ttttgttatt atgttatgag   1860
gttttaaaaa attgtttttg aatgtagaag atatatatta atattgtggg aaatataaga   1920
aaggataaga aattaagaaa ttataatgtt atttattat ataggttagt taattatgta   1980
ttttgtagag tagttgtata tattttttta agaaatgta tatagtgttg tatatggagt   2040
tttgtaattt ttttatattg attataattt aattaatttt tattaaagag ataaaagtga   2100
tgtttggtg tttatgtttt ttaggaatta ttaatagtta taattagttt tttagtaatt   2160
tttaattggg ttgtatttta aaaataatgt tttttatatt taatataaat gtatttttttt   2220
tttatatttg ggattaatat tgaaatttat gatttatta tattaaaatt taaatttttat   2280
tatattaata tttaaaattg tattagaggt tttatgattt ggtattatgg gtttttgtat   2340
```

```
tatttttttt ttaaattttt taatttgttt tattaaggtt tttggataat tttagagatt      2400
ttttgtgaag tttgaataaa attttttga dattttgata attgtattag ttttaggatt      2460
taattggaat agaattaaaa tttttaaaat aagtttttat a                          2501
```

<210> 16
<211> 2501
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 16

```
tataagagtt tgttttaagg attttaattt tattttaatt aagtttttaaa gttaatgtaa      60
ttattaaaat tttgaagaga ttttatttaa attttataaa aggttttttaa agttgtttag     120
aaattttggt gaaatagatt aggaaatttg gaaaggaaat aatgtggaga tttgtagtat     180
taaattatga gattttttaat ataattttaa atattaatgt aataaaattt aaattttggt     240
gtaataaaat tataaatttt aatattggtt ttaagtatag agaaaaagta tatttatgtt     300
gaatgtggaa aatattattt ttaaaatata gttgattaaa aaattgttgg ggaattgatt     360
ataattattg ataattttta agaaatatag atattaaaat attattttta ttttttttaat     420
agaaattggt taaattataa ttaatataag gaggttataa aattttatat ataatattgt     480
atatatttt ttggaaaaat atgtgtaatt gttttgtaaa atatatgatt aattagtttg     540
tgtgatggga taatattgta gtttttttaat tttttgtttt ttttgtatt tttatagta     600
ttgatgtata tttttttgtgt ttaaaagtaa tttttttaaag tttttataatg tggtaataaa     660
atattatgta tgttataaaa tttagaatat ggaaataaga agtttgtttt tttttttttt     720
ttattttggt ttggtttttt agatttttt ttttttattg gggtggggatt tttttgttgat     780
ttttttaagg tttttagttt ggtttgttaa ttgttgtata aaggtgttgt tttaggttag     840
agtttttata aagtgttggg tgagatttttt tttgtttgtt atgtttggtt gtggtaaagg     900
tgggaagggt tttggtaaag gtggtgttaa gtgttattgt aaagtattgt gtgataatat     960
ttagggtatt attaagttgg ttatttggtg ttttgtttgt tgtggtggtg tgaagtgtat    1020
ttttggttttt atttatgagg agatttgtgg ggtgttgaag gtgttttttgg agaatgtgat    1080
ttgggatgtt gtgatttata tagagtatgt taagtgtaag atggttattg ttatggatgt    1140
ggtttatgtg tttaagtgtt agggttgtat tttttatggt tttggtggtt gagtgttttt    1200
ttttattaat aaaaggtttt ttttaggggtt tttattttttt tagttgagga gttgtgatgt    1260
ttgtttgttt agttttttttt tattatttgt ttgtgtttgt tgagtttgtg ttgttattgg    1320
agtatgtgtt tttagttgtt gtaagtaggt attagttatt aattgttttt tagtaaataa    1380
aatatttaat ataatttgtt ttaggtgttt tagagtttat tgatatgggt tgttgtgtag    1440
attgtagtgt agtgttatta tggttttattg tagttttttg attagttgga attttgtgtt    1500
tattattatg taaggttaat tatttttttta aagattgggt ttttgtgtgt ttttttaggtt    1560
agttttgata ttttggttttt aattttttttg tttatgttta atgttggtat tatagtagtg    1620
agttattatg tttggttatg atattgttga ggttttaggg ttagttatat ttaagggggta    1680
attttttgtag tgtagtgggg aggaaagtta agtagttata ggttttttggt gttgtgaatt    1740
taattgttga atatagtaag aatttattat gtaataattt ttaattagtg gattgtaaat    1800
aagtttagtt tattggtttt ggtttaattt ttgagaaagg tgagaatttg aattttttgag    1860
tagaaatatg tagtttatta gtataataaa gtgattataa taaagattaa tttattttttt    1920
gttatttaat gtgggtagag ttttattgtt gagaattttt tagttttttga gtgtttttttt    1980
atttataatt gaagttgata agatggtatt aaaagtgaga ttttagtaa gtaaatttaa    2040
aggtttgaag gattagtgat taatgggaag tgttaataag ttatatttga ggttatttaa    2100
tttgagattt tgatttaatt ggtttaagga tattattatt tgtgggtta gatttgaaaa    2160
ataataattg ttttaaggaa ggaatttgtt ggtatttaaa taaaaaattt tttttttttga    2220
tttttatatg ttgttatttta gataaatttt tggtttatat ggttggattt ttttttgtta    2280
gttaaatttt gtgttatttt gtgaaaaaga attgagatta ataaaatggt agtgagaata    2340
tagggaaaga taaaaattat agtaagttta tttagttatg ttgtagttat tttttataagg    2400
ataggaagg attttgtttt ttatttatta tttaaagtgt gagtataaat tttaaaaata    2460
tatgatatta taggaataat gaagatgttt ttatttgtaa a                          2501
```

<210> 17
<211> 2251
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 17

```
ggattttgag ttgtttttag gtagttagga gtggtttttag tggtagttgt aaattttagt      60
agttgtttgt gttgtttgtg gttggggtgg agggtagtta gagttgggga ttaaggtttt     120
gtgttatttg tgtgtatagt tttatatttg aatgttgttt ttttgtagat gagattggtg     180
ggtattgtaa agttgggatt tgtttttgaa ggaaaaaaaa tagtgagtaa gaaatttagt     240
```

```
attattttt  attgatttat  tttgttgtat  ttttgtttt  ttaagttttt  gaaagttggt      300
aattttgatt  ttggtgttta  aaaattgata  gttattgaga  ttggttttga  gaagttgaag      360
atttggtagt  ttttagattg  agtaggataa  ggtgaaagta  ggttggaggt  gggtttagga      420
tatttgaggg  ttgattttgg  gggtttgtga  ggttgttatt  gttgttgttg  ttataggtga      480
gatggtgttg  ggttgatgtt  ggggttaatg  ggtagagaat  gtagggatgt  ggttttgtt       540
gaagagagtt  aagaagggaa  gagtgtgttt  tttaaattgt  ttttgtaatt  tgttttagt       600
gagtatttta  ttgatttaga  atttattgag  aatagtatta  gtgagttatt  ttttttttga      660
gatgggtttt  atttattttg  gtaatggagt  gagttggatt  gtggggagga  agaggaatgg       720
gaaaattagt  ttataaatat  taatgttagt  aagagtgtgt  tgttggtagg  atgtattgtg       780
agtttggaga  ttttggtggt  tgtagttggt  aagtggttat  aatttagaaa  gtaggattga       840
gttgtttttt  ttgtttattt  agtgtattgt  ttttgggtg  tgggtttaat  atttataag       900
tggtaatttt  tgtttatggt  agtttgtttt  ttttttatt  attttaat  ttagtttgt       960
attttaaggt  tgtgtattgt  tagttattat  tatgtttatt  tttggggtta  atttgtttgt     1020
tttttttgagt  tttgattggt  tgaatagttt  agtgattatt  ttggtggtga  tgtttatttt     1080
tggggtggtg  ggtaattggg  tggttattgt  ggtgttgtgt  aagttgtgta  aggagtagaa     1140
ggagatgatt  ttttatatgt  tggtatgtgg  gttggttgtt  attgatttgt  tgggtatttt     1200
gttggtgagt  ttggtgatta  ttgttatgta  tatgaagggt  taatggtttg  ggggttagtt     1260
gttgtgtgag  tatagtattt  ttattttgtt  ttttttagt  ttgtttggtt  ttagtattat     1320
ttgtgttatg  agtgttgagt  gttatttggt  tattaattat  gttatttt  atagttatta     1380
tgtggataag  tgattggtgg  gttttatgtt  ttttgtagtt  tatgtgttta  atgtgttttt     1440
ttgtgtgttg  tttaatatgg  gttttggtag  tttgtggttg  tagtatttag  atatttggtg     1500
ttttattgat  tggattatta  atgtgatggt  gtatgttgtt  tatttttata  tgtatgtggg     1560
ttttagtttt  tttttttatt  ttgttattgt  tttttgtaat  gtgttgtgt  gtggtgtgtt     1620
gttttgtatg  tattgttagt  ttatgtgttg  tattttgttg  ggtattgagt  agtattatgt     1680
ggttgtggtt  gttttggttg  ttttttgggg  ttattttgtt  gtttttttag  ttttgttgtg     1740
ttttagtgat  ttttggtgtt  gttggagttt  ttgttgtatt  gtgggtgttg  agatttagat     1800
ggttatttta  tttattgtta  tttttttggt  ggtgtttatt  tgttttattt  tgtttgtggt     1860
gagtgattgg  ggttggggtt  ttatttggtt  ttttttttgt  atttatttt  tgtgtttatt     1920
ttttgttttt  tgttttttt  ttgagttttt  ggtagtgaat  gtgttgtttt  taggttgggg     1980
ttgggatttt  tatattgttt  tttagagtag  gtttaatttt  ttttgaagtt  ttaattttaa     2040
tgagatttag  taggtgtttt  gttttttatat  tttttagttt  atgttttgg  aagtttgggt     2100
ttttttttt  attgagatag  tttttatttt  ttgttgtttg  atattggttg  agtttttaa      2160
gaaaatttt  tgttttttt  gttagatgtg  gaggggagtt  tgttgtagtg  tgatttagtt     2220
tatttttttg  tattgtgaat  tgtgaattgt  a                                      2251


<210> 18
<211> 2251
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 18

tgtagtttat  agtttatagt  atggaggaat  gggttaagtt  atattatagt  aggttttttt       60
ttatgtttaa  tagaggggt  gggggtttt  tttggaagat  ttggttagtg  ttaggtagta      120
aggggtaagg  gttgttttgg  tggagaaaga  aatttaggtt  tttgggaata  taaattgggg      180
gatgtagggg  taaagtattt  gttaaatttt  gttagggttg  ggattttaaa  gagggttggg      240
tttgttttga  gaaatagtgt  gggaattta  gttttgattt  aaaggtgata  tgtttattgt      300
taaggattta  gagggaaagt  agggagtggg  gaatggatgt  gggaggtgga  tgtgagaaaa      360
aggttgagta  gggtttagt  tttggttatt  tattatgagt  gggatgggagt  agatgagtat      420
tattagggag  gtggtaatga  gtaagatgat  tatttgtgtt  ttggtgtttg  tgatgtggtg      480
gaagtttttgg  tggtgttgaa  agttgttgag  gtgtggtaag  gttggggagg  tagtggggtg      540
gttttgggag  gtaattgagg  tggttgtggt  tgtgtggtgt  tgtttggtgt  ttagtgaggt      600
gtggtgtatg  aattggtggt  gtatgtggag  tagtgtgttg  tatataagta  tgttgtagag      660
gatggtggtg  agaatgagga  aggagttgaa  gtttgtgtat  atgtaggagt  aggtggtgtg      720
tgttgttatg  ttggtggttt  agttgatgaa  gtattaggtg  tttgggtatt  gtagttgtga      780
gttattgaga  tttatgttgg  gtagtgtgta  aaagagtatg  ttggatgtat  agattgtaaa      840
gagtgtgagg  tttgttaatt  gtttgtttat  gtagtggttg  tagaaatagg  tatggttgat      900
ggttaggtag  tgtttgatat  ttatggtgta  gatgatgttg  aggttggata  ggttgaagaa      960
gagtagaatg  aaggtgttgt  atttgtatag  tggttggttt  ttgggttatt  ggttttttat     1020
gtatgtggtg  atggttattg  ggtttattaa  taaagtgttt  aataggttgg  tgatagttag     1080
tttatatatt  agtgtgtaga  aggttgtttt  tttttgtttt  ttgtgtgatt  tgtatagtat     1140
tatgatggtt  attaggttgt  ttattatttt  gaagatgaat  attattgttg  ggatggttat     1200
tgggttgttt  agttggttgg  ggtttaagga  ggtggatgaa  ttgatttgg  gagtggatat     1260
gatagtggtt  ggtggtgtgt  agttttggag  tgtaaggttg  ggttggggga  tggtagaaga     1320
gagataaagt  tgtttgtgagt  ggaaattatt  atttgtaagg  tgttagattt  gtgtttgaga     1380
aatgatatat  tgataagata  aggggagtaa  tttgatttta  tttttggat  tgtagttatt     1440
tattaattgt  ggttattaaa  atttttaggt  ttgtgtatatg  tttttgttaat  agtatatttt     1500
tgttgatatt  aatatttata  aattgatttt  tttattttt  tttttttta  taatttaatt     1560
```

```
tattttattg ttaagatgaa taagatttat tttaagggaa aagtagtttg ttagtgttat      1620
ttttgataga ttttgaatta ataaaatgtt tattgaaaat aagttataaa agtaatttgg      1680
agagtgtgtt tttttttttt tggttttttt tggtgagggt tgtatttttg tgtttttat      1740
ttgttgattt taatgttagt ttaatgttat tttatttgta gtggtagtag tggtggtagt      1800
tttatgagtt tttagggtta gtttttagat gttttggatt tgtttttaat ttgttttttat     1860
tttgttttgt ttagtttgga aattgttaaa ttttttggttt tttaaagttg gttttagtgg     1920
ttgttgattt ttggatattg aggttagagt tgttagtttt taaaaatttg ggaaataaga      1980
ggtgtagtgg gatgggttag tgaagaatgg tgttggattt tttatttgtt attttttttt      2040
ttttaaagat gagtttttagt tttgtagtgt ttgttggttt tgttgtgggg aggatagtgt     2100
ttaggtgtga ggttgtgtgt gtaggtggtg tggagttttg gtttttagtt ttggttgttt      2160
tttgttttag ttatgggtag tatgggtggt tattggagtt tgtggttgtt gttgaggttg      2220
tttttggtta tttggggggtg gtttggggtt t                                    2251
```

<210> 19
<211> 2586
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 19

```
tttgtgtggg tgtgggagta gttttgttgg gtgtttgtag ttgtgggagt taagtttttt       60
ttttaggtgt aggtataaaa gtttatggtt tttgaataat gtggggtaga ggtttttta       120
agtaatgttt aattggttgt ttttaattaa ggaaagagag gtttttagtt ttatggtaat      180
ttaagtaggg tagttttagg ttaaaggtat tttagaataa taagattatt ttaagaaatg      240
gaatgtttta ttggatattt gaataggttt tttgttattg gaattggtgt gtattgtatt      300
ttaattagta tttttgtgtg gagggaggtg atttagttta ggaaagttaa ttatagaaag      360
aggtgatttt tgaaaggatt gtttttagtgt tattagaata tatgtgatta tattaaaagt     420
ttaggtggat atttgtagtt agtttggatt tggataattt aatattgttg gtagaattga      480
agggaataag ttattttaat tttatttttt gtatgtgtag tgttgagtga gttgggggtg      540
ggaggatagt ggtttgttta ttgtttttttt ttaaaatttg agatttgaaa atatggaggt     600
tttatttgtt tttttagttt ttgattgtta ataaaaaaat aaattttgtt ggttatattt      660
ttttttttatt ttaaaatagt aattttatgg tttgtattaa gttttttaga agtttatttt     720
atttgttttg ggttagggag ggaataatgt taggaaaagt tattggtgtt ttttatttt       780
tgtttttttt agggtgtagg atgtgtgggt tggtgggttt gtgattttgg aatgtttttt      840
gttattttt tgtgtgaatt tgaaaggatt gggaggtgtt gagagtagag tttagggttg       900
gtgtattttg tggtgttgag tgggtggtgt gtttgggtgt ttaggttggg ggatttgtag      960
ttgttttatt gtggagggga aaatatgtag ttggagggtg tgtgttgtgt gggttgtgat     1020
ttgttatttt attggttatt ttggggtttt tttaagtttt taggtagggt tgtgagagtt     1080
ttttagagtt gaagttttgg aggttgattt gtgggtttgg ttgttatggg aatttggttg     1140
gtttaaagaa gttttttttt tgggtatttg gaattttagt ttagtgtggg gtattgggga     1200
agtggtgttg ggggggttggg ttgggggatt ttagttggta gtttttggaga gggtttattt    1260
ttggggttgt tgggtgaggt tggtatgatt tttggtttta aaaggaaagt ttttgttttt      1320
tgttttggtg tgagaagtta aagatttatt ttgagagtgg agagagaaat gttattggta     1380
atgttttttt tggaaagttt gagaggggtt ttttggatat attatttagt gttttaaat      1440
tagagaagta gtttttttttt ggtgtttggg tttagaagtt gttatttatt tagtttatgg     1500
tttgaaatta gtatgggaag tgttggggta aggttttgtt ggagattaga ggtttgtttg     1560
ttgggaggag ttttggggg atggggattt tatttttttg tttgtttttgg ttttttattttg    1620
ggatgttttt gtaggagttt agaaagatga tttattatat ggttttggga tagagtagtg     1680
tgtttaattt tgagggaatt ttgtgtgttt tttgaggtt ttagtttttt aaggtattgt       1740
tgtttgtttt tttttttttta gattgaaatt ggggaagagt gtgggtgttt ttttgttttg    1800
atgagtttgt tttttttaaat gtttatttttg gttgtattag agtatttggg aaatttttgaa  1860
aggtgtttag gtttttatata gtagtgtttt tttatttagt ttttgttttt gggtttttttt    1920
aagagagttt ttattttatg ttttggtttt ttttttgatgt tgggtttttg aggtaggtat    1980
ggagtttttt tgaaagtagt tgtttatttg tgttttttttg gtgtaaagtt agagtttatt     2040
ttgttggggg aagggggaggt agaaaagatt atagttggga aagtgtgttt tttgttttgt    2100
tttttaaaata tgttttaaga ttgttattgt gattgttagg agagttatta atgtttaggg    2160
gttataaagg aattttttgaa tttttggttt tttttaaatt tttaggtttt taaaatttta     2220
gtgggggttt tttggggttg ggatttaggt tggtattgtt gggaggattt tgtttagtat     2280
tttttttatta atattttatg aaggtaggtt tttgttttt ttggagtttt tttttttgga      2340
atgtttttaa attttggtta atttattttt ttgtgagtta ttttaggggtt ttgtggtttg     2400
ggaagagatg tgttaatttt gtgggtttgt gtgtagtttt tagttgtaaa gtgttgtaag     2460
tgattttttt gatggttttt tttgattgaa gagtttggga attaaagaga aaaaaaataa     2520
ttttatttttt aaaagaataa gttattattg tggtgatatt gtggtggagg attttggtga    2580
tggggt                                                                 2586
```

<210> 20
<211> 2586
<212> DNA

<210> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 20

```
attttattgt taaagttttt tgttatagta ttgttgtagt gatgatttgt ttttttgaaa    60
ataaagttat tttttttttt tttggttttt gagttttttg gttggaaagg gttgttaggg   120
gggttatttg tagtattttg tggttaagga ttgtgtgtag atttgtggag ttgatgtgtt   180
ttttttttggg ttatagagtt ttaggatggt ttatagggga gtgagttagt tagagtttgg   240
gaatattttg agaaaagagg ttttagagaa ggtaggagtt tgttttttgtg aaatattaat   300
aaagggatgt taggtgaggt tttttttagtg gtgttagttt gaattttagt tttaagagat   360
ttttattagg gttttaggaa tttggggggtt tgggaagggt tgagggttta gaaattttttt   420
tatagttttt agatgttgat agtttttttta ataattgtga tgatagtttt gaggtatgtt   480
ttaggaataa ggtgaaaagt gtatttttttt aattgtgatt tttttttattt ttttttttttt   540
taataaagta aattttaatt ttatattaaa ggggtataga taggtaattg tttttagagg   600
gattttgtgt ttatttttggg gatttgatat tgaagaaaga ttgaggtatg aggtagagat   660
ttttttgaaa aaatttaaag atagaggttg agtagggaga tgttgttgtg tgaggtttgg   720
gtatttttta gagtttttta gatgtttttgg tgtagttgaa gtaggtgttt ggggaggtga   780
atttattggg gtaaagaagt gtttatattt ttttttagtt ttggtttagg gaaagaagaa   840
tggatggtgg tgtttttggaa agttagggtt ttagagaggt gtataaagtt ttttttaaagt   900
tgggtgtgtt gttttgtttt gggattatgt agtggattgt ttttttgggt ttttatggag   960
gtgttttagg tgggaattag agtaagtagg agaatggggt tttttatttttt taggggtttt  1020
ttttgatagg taggtttttta gtttttgatg aagtttttgtt ttggtgttta ttatgttgat  1080
tttgaattat gggttgagtg agtggtgatt tttgagttta ggtattaaag aaaaattatt  1140
ttttttggttt ggggggtatta ggtagtgtgt ttagaagatt tttttttgaat tttttaaaga  1200
aaatgttatt aataatattt ttttttttttgt ttttaaaata agttttttggt ttttttgtgtt  1260
agaataagaa gtagaaattt ttttttttgga gttaagaatt gtgttggttt tatttagtga  1320
ttttaagggt aggttttttt tggagttgtt ggttgaggtt ttttaattta gtttttttagt  1380
gttatttttt tgatgttttta tattaaattg gaatttttagg tgtttggaag aaaggttttt  1440
ttggattaat tgggtttttta tagtagttag atttataggt tagttttttgg ggtttttagtt  1500
ttaggggatt tttatagttt tatttagagg tttggggaga ttttaggatg attgatgggg  1560
taatggatta taattttgtat ggtgtatgtt ttttagttat gtatttttttt ttttgtggta  1620
gggtgattat aggtttttttg gtttgagtgt ttgggtgtgt tgtttatttta gtattgtaga  1680
gtgtattagt tttgaatttt gttttttaata ttttttttagtt ttttttaagtt tgtgtaaggg  1740
gatggtagga agtgttttgg gattataggt ttgttggttt gtatattttg tattttggaa  1800
ggggtgagga tggaagagta ttggtgattt ttttttagtat tgttttttttt ttggtttaga  1860
gtaaatgaga taaattttttg aagggtttaa tataagttat agggttgtta tttttggaatg  1920
aggagaaaat gtagttagtg ggatttgttt ttttgttggt aattaagagt tgggaaaatg  1980
aatgggattt ttatattttt agattttaga ttttaaaagg gggtaataaa tgaattgttg  2040
tttttttttatt tttaatttat ttagtattat gtgtatagggg gatggggttg gggtaatttg  2100
tttttttttag ttttgttagt aatgttgaat tgtttaaatt tgagttggtt gtgggtgttt  2160
gtttgggttt ttggtgtggt tatatgtatt ttaatagtgt taagataatt tttttggagg  2220
ttatttttttt ttgtagttga tttttttaga ttgggttgtt ttttttttata taagagtatt  2280
ggttgaagta tagtatatat taattttaat gatagagaat ttgtttgggt gtttagtgag  2340
atatttttatt ttttagaatg attttattat tttaaagtat ttttagtttg aagttgtttt  2400
gtttgggttg ttatagagtt ggaagttttt tttttttttaa ttagaagtgg ttaattagat  2460
gttgtttgga aaaattttttg ttttgtattg tttaagagtt ataaatttttt atgtttgtat  2520
ttggggaggg ggtttgattt ttgtggttgt gagtgtttag tggggttatt tttgtgtttg  2580
tgtggg                                                              2586
```

<210> 21
<211> 2308
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 21

```
ttaggtagga ggtagttagt tggtgatagc gcgcgtgatt atagattttt gatttttata    60
gttttagtag ttgtttgtag gtttttttacg ggatcgtagt ttgggagggt aggaggtttc   120
gtatttgcgg tcggtttgat ttgtttttta ttttttttgtt tgttttttttc gtgttgggcg   180
tcgtcgcgat acgtttcgcg gatatttagg agttttttta gttttgcga gcgcggattc   240
gggagagagt ttttatttaa cgttgaggtt attggtagtat agatttttttc gggttatttt   300
gcgaagtgtt tggtttttttt tattttttagt tttgtagtag taataaacgg gaaggaagtg   360
ggtgacgagg ataggtagag atttttttta cggggtcgtt cgtattttat ttgtcgtcgc   420
gggggcgggcg ttgcggtcgg tgtttagttg tagattcgta ggcgatcgta tttttggcgg   480
tttttagtttt aagtaggtag cggtcgattc gtttgtttcg aggggtttagg gttgatttag   540
```

```
gtgagtttcg gcggggaggg cgggcgtttc ggttttttcg tttcggaatt ttgtatcgtt    600
tagagtcgtt acgtttttcg gttgtaggtt tcgtatttgt aaaatgggtc gatgaatttt    660
aagggtcggt ttagatttac ggattgtagt tgtagttggg ttagaagggg tggttagggg    720
agtcgatttt attttttttt ttttttttacg ttaattagga gtttttattt gtttgggatt    780
tttttgggggt atttatgggt ttatagtcgt agaattatag tattgggtgg ggttttagtt    840
ttggggttaa gtttttgtga ggttgttat tatttttttt tgttgttttg ttttttgtt    900
tgtttgtttg tttggttttt tttttttttga gatagagttt cgtctttta tttaggtcgg    960
attgtagtgg cgcgatttcg gtttattgta agttttgttt ttcgggtttt cgttattttt   1020
ttgtttttagt attttcgagt agttgggatt acgggcgtcg ttatcgcgtt cggttaattt   1080
tttgtatttt ttagtagaga cggggtttta tcgtgttagt taggatggtt tcgattttt   1140
gatttcgtga tttatttatt ttggttttttt aaagtgttgg gattataggc gtgagttatc   1200
gcgttcggtt tgttttttgtt tttttttttt ttgagatagt tttattttgg ttgtttaggt   1260
tggagtgtag tggcgtaatt ttggtttacg taattttcgt tttcgggta taagcgattt   1320
ttttgtttta gtttttttag tagttgtgat tataggtacg tattattacg ttcggttaat   1380
tttgtatttt tagtagagac ggggtttttt tatgttggtt aggttggtt ttaatgttcg   1440
attttaggtg atttgtttat tttagttttt taaagttttg cgattatagg cgtgagttat   1500
tatgtttggt tagtttttg ttttttttaaa aatttattat ttagtcggtg tggtgtttta   1560
tatttgtaat tttggtatttt tgggaggtcg aggtgagcgg attatttgag gttaggtgtt   1620
ttagattagt ttgggtaata tggtaaaatt ttgttttttat taaaaagata aaaattagtt   1680
ggaggtgttt ttttgaattt aggaggtaga ggttgtagtg aattgagatt atgttattgt   1740
attttaattt aggaggagga ggttgtattg agttaatatt atgttattat attttagttt   1800
gggcgataga gtaagatttt gtttttttta aaaaaaaat agaaaaaaa aaaagaaag   1860
aaatttgaag taattatgat taaatgttaa tatttgtttt tttacgtggt gggtataggc   1920
gtgttttgga agtttttatg ttttttgtat tttgagatta tataaaatta tgtttattat   1980
tttattgtgt tatatgattt ttgtagaaaa gttggtaaat tttggaatgt acgaagaaaa   2040
agattataat agttagtaat attttttgttt agagagatag tatttattaa tttttttttg   2100
tttatgtttg ttttcggaaa aatgggatat tattggtatt attttataat tttttaatttt   2160
gtattaatat tttttatatt ttttaaattg aaaaagttgg ttaagtatgg tggtttaggt   2220
ttgtaaatttt agttaatgtg ggaggatcgt ttgagttcgg gagtttagga atagtttggg   2280
tatcgtggcg agattatatt gttataaa                                     2308
```

<210> 22
<211> 2308
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 22

```
tttgtagtaa tgtggtttcg ttacgatgtt taggttgttt ttgaatttttc gggtttaagc     60
gatttttta tattggttgg gtttataggt ttgagttatt atgttggtt agtttttta    120
atttaaaaaa tatagaaagt gttaatataa attaaaaagt tataaaatag tgttaatagt    180
attttatttt ttcgaagata aatataggta gagaaaaatt aatggatatt gttttttttgg    240
gtaggggtat tattggttat tataatttttt ttttcgtgt attttagaat ttattaattt    300
ttttataaag attatgtagt ataataagat aataaatata attttgtgtg atttttaaaat    360
ataggaagta tagaagtttt taaaatacgt ttgtatttat tacgtgaaga aatagatgtt    420
aatatttggt tataattgtt ttagattttt tttttttttt tttttttttgt tttttttttg    480
gggggatag agttttgttt tgtcgtttag gttggaatgt agtggtatga tattggttta    540
gtgtaattttt tttttttttgg gttggagtgt agtggtatga ttttagttta ttgtaatttt    600
tgtttttttgg gtttaagaga gtatttttag ttaatttttg ttttttttagt agagataagg    660
ttttgtttatg ttgtttaggt tggttggga tatttgattt taagtgattc gtttattttcg    720
gtttttttaaa gtgttaggat tataggtgtg agatattata tcggttaaat aataagtttt    780
taaaaaaata ggagattggt taggtatggt gatttacgtt tgtaatcgta gaatttttggg    840
aggttgaggt gggtagatta tttgagatcg ggtattggag attagtttgg ttaatatgga    900
gaaatttcgt ttttattaaa aatataaaat tagtcgggcg tggtggtgcg tgtttgtaat    960
tatagttatt agggaggttg aggtaggaga atcgtttgta ttcgggaggc ggaggttgcg   1020
tgagttaaga ttgcgttatt gtattttagt ttgggtaatt agaataaaat tgttttaaaa   1080
aaaaaaaaaa taaaaatagg tcgggcgcgg tggtttacgt ttgtaatttt agtattttgg   1140
aaggttaagg tgggtggatt acgaggttag gagatcgaaa ttattttggt taatacggtg   1200
aaatttcgtt tttattaaaa aatataaaaa attagtcggg cgcggtggcg gcgttcgtag   1260
ttttagttat tcggggggtgt tgaggtagga gaatggcgag aattcgggag gtagagtttg   1320
tagtgagtcg agatcgcgtt attgtagttc ggtttgggtg aaagagcgag attttgtttt   1380
aaaaaaaaaa aaattaaata aataaataaa taaaaaaata aaataatagg agaagatgat   1440
gatagatttt ataagagttt ggttttaggg ttgaaatttt atttagtgtt gtggttttgc   1500
ggttgtgggt ttatgggtgt ttttaaaggg ttttaggtag atggaggttt ttggttggcg   1560
tggaggaagg aaaaggatga aatcggtttt tttggttatt tttttttagtt tagttggtagt   1620
tgtagttcgt gggtttgagt cgattttttgg ggtttatcgg tttattttat agatgcggag   1680
tttatagtcg ggaggcgtgg cggtttttggg cggtgtaggg tttcggggcg gggaggtcga   1740
ggcgttcgtt tttttcgtcg gggtttatttt gggttagttt tgagtttttcg agtagggcgg   1800
```

```
atcggtcgtt gtttgtttaa agttggggtc gttaggagtg cgatcgtttg cggatttgta      1860
gttgggtatc ggtcgtaacg ttcgtttcgc gacggtaggt gagatgcggg cggtttcgtg      1920
ggaggggttt ttgtttgttt tcgttatttta tttttttttc gtttgttgtt gttgtagggt      1980
tgggaatggg ggggattaag tatttcgtag ggtgattcgg ggagatttgt gagttagtgg      2040
ttttagcgtt gaatgagagt ttttttttcgg gttcgcgttc gtagggattg aagggatttt      2100
tgaatgttcg cgggacgtgt cgcggcggcg tttagtacgg aggaggtagg tagggaggtg      2160
aggagtaggt tagatcggtc gtaggtgtag ggttttttgt ttttttaggt tgcggtttcg      2220
tggagggttt gtagatagtt gttgagatta tgagggttag aggtttgtgg ttacgcgcgt      2280
tattattagt tagttgtttt ttgtttgg                                        2308


<210> 23
<211> 2553
<212> DNA
<213> Artificial Sequence


<220>
<223> chemically treated genomic DNA (Homo sapiens)


<400> 23

ttggttattt ttgtttttttt ttagattttt agagttttta gtgtagttat agaggatgtt        60
gttggtttta gttttaagaa gacgtcgttt tttttagagg gttaagtaag tgggaatttt       120
ttttttttatt tgtttgggt tttaggtagg gtttttggtg taaggtttgg ggttggaagt       180
cgatttattt aggtttaggt tttggggtag aattgaaatt ttttggttat tgtcggttgt       240
agtttgggag taggttattg .ttaaagttgt gggttttttt aggatagttt ttttatgagg       300
tcggttttttt atttgttgtt tttttatatt tggtggttag ggatgtggtt ttgggtagaa       360
cgatgatttt ttattttttgt tattatgaa gttatcgttg tttttttagtt tagttagtta       420
tttgggttgt agagtatttt ttttatgttt ttcgggtgtt ttttttttttt tttgtttttag       480
tttggtttttg ttttatttttg ttttttaggga ggggtatttt ggagtggggt tagggtatgg       540
tttttttttcg agggagtttt tttttggttg ttttttagggt agttttgtat agttttagta       600
tttggcgtat ttttttttgat attttttttta gggatagtta ggtattttgt gtggggtatt       660
taagagagtt aggttcgtta gttttttagtt tttgttagaa tgtaggtttg aggggtgagg       720
ggcggggtag gggtagggat aggaatttcg gcgtgttttt tattcgtaaa ggtttattga       780
ggtttcgagt tttagttatt gagttattaa gttagtttgg gttaggtttg ggtgttttgt       840
ttgtaatgga ggtagagacg gggtttcggg gtagtttttga ggatgttggg tgtatagcgg       900
gggtttcgtc ggtaggaatt atttatgttt tttttttgggt taagtttttgt ggatgtttag       960
tttggggtcg cggggagttg gtaggttagt ggtagatatt ggtgggtaga tttagtgttt      1020
ggtagaatag gtattaagga agtggtgatc ggagggaagt taagtgtatt taaatttttcg      1080
ggtgagttat tatcgtcggg ttttttttatag ttgttgaaag tgagtaatag tgatgaaggt      1140
ttgtgagttt ttgcgtgagc gagtgaattgg attagtagta gtttttaggt tgtggaagag      1200
cgtttttttt tcgggatggg gatatttggt tatagtaatt tttaatttttt tatttatttta      1260
tcgtttattg tagaggtatg cggggggtttt gtttttttgta ggtaggagtg aggggtattt      1320
ttgtgatgtg gtatttttgt gatcgaggtt atgtgtgatc ggtgtaaggg taggaagcga      1380
gttattggtt tgtattaggc gtgggggttt ttgcgagggt aggatttaaa gtcggtttgg      1440
ttttttcggtt gtagtatttt ttttttttttc gaattaggtt agagttttgg gacgggaggt      1500
gttttgtaga ttatttttttt tattaattttt cgttttttcgt tttatttttcg cggtgattcg      1560
gtgaattgtc ggttttttgt tgtgtatcga gtggggtagt gattttgacg tggcgttttt      1620
tgtcgttttt gttatcgtta ttattttcgg tggtttagtt ttcgtatttt ttatttttat      1680
ggaggaatgt attaggtttt tttttttttgga tgtatttttt atttatatgt ttttaaattt      1740
tggtatttttt tgtttttttttt ttatttttat ttttttttttt aggttttttag ataaaggggga      1800
agtggttgga tttttttaaa gggatagtgt tttattagtt tattgttgaa ttttttttttt      1860
taatttttagt ttttttagtta tagttaatta gtattagtag atagtttatg agtgatattt      1920
atgtaggttt taggttgtgg agagtttttt gggtaggaaa tagtttttaa ggttttttat      1980
tttattttagg ttttagtttt ttttatttgt ttttttttttta gattgtggtt ttttggagtt      2040
tggttttttttt gttttgtgt gatcgatata tagtatttaa atagtggtag agcgggacgg      2100
attttttttagt ttgtttttttg tgtgggtttg tattttgatt tagatatgtt tttttatagt      2160
aggatttagg ggggtatatg tgtgtttgcg ggttttattgg ggtattcgta tttggtttat      2220
tttattttttt agagagaggg ttttgttgtg ttatttagtt ggagtgtagt ggtgtaatta      2280
tagtatattg tagttttttaa tttttgggtt taagcgattt ttttttttttta gtttttttttag      2340
tagttgggag tataggattt attgtatttt ggttaatttt ttaataatttt tttaagagat      2400
ggggtttttat tgtgttgttt aggttggttt taaattttttg gttttaagtg attttttttat      2460
tttcgttttt tgaagtgttg agattatagg tatgagttat tatgtttatt ttagattgat      2520
atttttatat ttgtttattt tggttgggta ggg                                   2553


<210> 24
<211> 2553
<212> DNA .
<213> Artificial Sequence


<220>
<223> chemically treated genomic DNA (Homo sapiens)
```

<400> 24

```
ttttgtttag ttaggatgaa taaatataga aatgttagtt tgggatgggt atggtggttt      60
atgtttgtaa ttttagtatt ttaggaggcg aaggtgggag gattatttga ggttagaggt     120
ttgaggttag tttgggtaat atagtaagat tttatttttt aaaaaattat taaaaaatta     180
gttaggatat agtgggtttt gtatttttag ttattaggga ggttggggag ggaggatcgt     240
ttgagtttag gagttgaagg ttgtagtgtg ttatgattat attattgtat tttagttggg     300
tgatatagta agattttttt tttaaaaaat aaaataaatt aaatgcgggt gttttagtga     360
attcgtaggt atatatgtgt ttttttgggt tttgttgtgg gggggtatgt ttgggttagg     420
gtatagattt atatagagaa taggttaggg ggttcgtttc gttttgttat tgtttgggtg     480
ttatgtgtcg gttatatagg gatagaagaa ttaggtttta aagggttata atttaggaga     540
gaggtaggta ggaaagattg ggatttggat gagatgaggg attttaaggg ttgttttta      600
tttaggaaat tttttatagt ttgggggtttg tatggtgatt atttatgggt tgtttgttaa     660
tgttaattaa ttgtaattag ggaattgggg ttgaggaggg gagtttagta gtaagttggt     720
gggatattgt ttttttaaga ggatttagtt attttttttt tgtttgggag tttaggggaa     780
ggggtgggag taaagggggga gtagaaaatg ttagggtttg gaagtatgtg ggtgaggggt     840
gtatttagga agggaggttt ggtgtatttt tttatggggg tggggaatgc ggaggttggg     900
ttatcggagg tggtggcggt ggtaggggcg gtaggagacg ttacgttagg gttattgttt     960
tattcggtgt atagtagggg gtcggtagtt tatcggatta tcgcggggggt ggggcggagg    1020
acggaagttg gtgagggggg tggtttatag ggtatttttc gttttagggt tttaatttaa    1080
ttcgaaaggg aaaggagtgt tgtagtcggg aggttaggtc gattttgggt tttgttttcg    1140
tagaagtttt tacgtttggt gtagattaat gattcgtttt ttgtttttat atcgattata    1200
tatgatttcg gttataggg gtgttatatta taggagtgtt ttttatttt gtttgtagga     1260
agtagggttt tcgtatattt ttgtagtggg cggtgggtgg gtgggggatt aggaattgtt    1320
gtaattaagt gtttttattt cggggaggga acgttttttt ataatttgga aattgttatt    1380
ggtttatta ttcgtttacg tagaaattta caaattttta ttattgttgt ttattttag     1440
tagttgtgaa agattcggcg gtgatgattt attcgagggt ttgagtgtat ttggtttttt    1500
ttcggttatt attttttga tgtttgtttt attagatatt aggtttgttt attagtgttt     1560
gttattgatt tgttagtttt tcgcggtttt aggttgggta tttataaagt ttggtttagg    1620
agagagtata agtgatttt gtcggcgagg ttttcgttgt gtatttagta tttttagaat     1680
tgtttcgaga tttcgttttt gttttttattg tagatagggt atttaggttt ggtttaggtt    1740
gatttggtgg tttagtggtt ggggttcggg gtttagtga atttttgcgg atggagagta    1800
cgtcggagtt tttgtttttg tttttgtttc gttttttatt ttttaggttt gtattttggt    1860
aggagttaga ggttgacggg tttggttttt ttgagtgttt tatatagagt gtttgattgt    1920
ttttaagaag gatgttaagg gaggtgcgtt aggtattgag gttgtgtaga gttgtttttgg    1980
ggatagttag agaggaattt tttcggggga gagttatgtt ttggtttttat tttagggtat    2040
ttttttttga gagtagggta ggataaagtt aggttggggt aggagaaggg ggaggtattc    2100
ggagggtatg aaagggggtgt tttgtagttt aggtggttgg ttgggttggg agatagcggt    2160
ggtttttata atgataggag tggagaatta tcgtttttat tagggttata ttttttggtta    2220
ttaggtgtga agaaatagta ggtggaggat cggtttttatg gggagattgt tttggaagga    2280
tttatagttt tggtagtggt ttgtttttag gttgtagtcg atagtaatta aggagttta     2340
gttttgtttt agagtttgga tttaggtggg tcggtttttta gttttaggtt ttatattagg    2400
ggttttgttt ggagtttagg ataagtaggg aggggggattt ttatttattt agttttttgg    2460
aggaagcggc gttttttttgg ggttgaagtt aatagtattt tttgtggtta tattgggggt    2520
tttggaggtt taagggaggg tagggggtggt tag                                2553
```

<210> 25
<211> 2308
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 25

```
ttaggtagga ggtagttagt tggtgatagt gtgtgtgatt atagatttt gattttttata     60
gttttagtag ttgtttgtag gttttttatg ggattgtagt ttgggagggt aggaggtttt    120
gtatttgtgg ttggtttgat ttgtttttta tttttttgtt tgttttttt gtgttgggtg     180
ttgttgtgat atgtttttgtg gatatttagg agttttttta gttttgtga gtgtggattt     240
gggagagagt ttttatttaa tgttgaggtt attggtttat agattttttt gggttatttt     300
gtgaagtgtt tggtttttttt tattttttagt tttgtagtag taataaatgg gaaggaagtg     360
ggtgatgagg ataggtagag attttttta tggggttgtt tgtatttttat ttgttgttgt     420
ggggtgggtg ttgtggttgg tgtttagttg tagatttgta ggtgattgta tttttggtgg     480
ttttagtttt aagtaggtag tggttgattt gttttgtttg agggtttagg gttgatttag     540
gtgagttttg gtggggaggg tgggtgtttt ggttttttg ttttggaatt ttgtattgtt     600
tagagttgtt atgtttttttg gttgtaggtt ttgtattgt aaaatggggt gatgaatttt     660
aagggttggt ttagatttat ggattgtagt tgtagttggg ttagaagggg tggttagggg     720
agttgatttt attttttttt tttttttatg ttaattagga gttttttattt gtttgggatt     780
ttttggggggt atttatgggt ttatagttgt agaattatag tattgggtgg ggttttagtt     840
```

```
ttggggttaa gttttttgtga ggtttgttat tattttttttt tgttgttttg tttttttgtt      900
tgtttgtttg tttggttttt tcttttttttga gatagagttt tgtttttttta tttaggttgg      960
attgtagtgg tgtgatttttg gtttattgta agtttgtttt tttgggtttt tgttattttt     1020
ttgtttttagt atttttgagt agttgggatt atgggtgtttg ttattgtgtt tggttaattt     1080
tttgtatttt ttagtagaga tggggtttta ttgtgttagt taggatggtt ttgattttt      1140
gattttgtga tttatttatt ttggttttttt aaagtgttgg gattataggt gtgagttatt     1200
gtgtttggtt tgtttttgtt ttttttttttt ttgagatagt tttattttgg ttgtttaggt     1260
tggagtgtag tggtgtaatt ttggttttatg taattttttgt ttttgggta taagtgattt      1320
ttttgttttta gttttttttag tagttgtgat tataggtatg tattattatg tttggttaat     1380
tttgtatttt tagtagagat ggggtttttt tatgttggtt aggttggttt ttaatgtttg     1440
attttaggtg atttgtttat tttagttttt taaagtttttg tgattatagg tgtgagttat     1500
tatgtttggt tagttttttg ttttttttaaa aatttattat ttagttggtg tggtgtttta     1560
tatttgtaat tttggttattt tggggaggttg aggtgagtgg attatttgag gttaggtgtt     1620
ttagattagt ttgggtaata tggtaaaatt ttgttttttat taaaaagata aaaattagtt      1680
ggaggtgttt ttttgaatttt aggaggtaga ggttgtagtg aattgagatt atgttattgt     1740
atttttaattt aggaggagga ggttgtattg agttaatatt atgttattat atttttagttt     1800
gggtgataga gtaagatttt gtttttttttta aaaaaaaat agaaaaaaaa aaaagaaag      1860
aaatttgaag taattatgat taaatgttaa tatttgtttt tttatgtggt gggtataggt     1920
gtgttttgga agttttttatg ttttttgtat tttgagatta tataaaatta tgtttattat     1980
tttattgtgt tatatgatttt ttgtagaaaa gttggtaaat tttggaatgt atgaagaaaa     2040
agattataat agttagtaat attttttgttt agagagatag tatttattaa tttttttttg     2100
tttatgtttg ttttttggaaa aatgggatat tattggtatt atttttataat tttttaatttt     2160
gtattaatat tttttttatatt ttttaaaattg aaaaagttgg ttaagtatgg tggtttaggt     2220
ttgtaaatttt agttaatgtg ggaggattgt ttgagtttgg gagtttagga atagtttggg     2280
tattgtggtg agattatatt gttataaa                                         2308
```

<210> 26
<211> 2308
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 26

```
tttgtagtaa tgtggtttttg ttatgatgtt taggttgttt ttgaattttt gggtttaagt       60
gatttttttta tattggttgg gtttataggt ttgagttatt atgttggtt agttttttta      120
atttaaaaaa tatagaaagt gttaatataa attaaaaagt tataaaatag tgttaatagt      180
attttattttt tttgaagata aatataggta gagaaaaatt aatggatatt gttttttttgg      240
gtaggggtat tattggttat tataattttt ttttttgtgt attttagaat ttattaatttt      300
ttttataaag attatgtagt ataataagat aataaatata attttgtgtg atttttaaaat      360
ataggaagta tagaagtttt taaaatatgt ttgtatttat tatgtgaaga aatagatgtt      420
aatatttggt tataattgtt ttagatttttt ttttttttttt tttttttttgt tttttttttg      480
ggggggatag agtttgtttt tgttgtttag gttggaatgt agtggtatga tattggttta      540
gtgtaatttt tttttttttgg gttggagtgt agtggtatga ttttagttta ttgtaatttt      600
tgttttttgg gtttaagaga gtattttag ttaattttttg ttttttttagt agagataagg      660
ttttgttatg ttgtttaggt tggttggaa tatttgattt taagtgattt gtttattttg      720
gtttttttaaa gtgttaggat tataggtgtg agatattata ttggttaaat aataagtttt      780
taaaaaaata ggagattggt taggtatggt gatttatgtt tgtaattgta gaatttttggg      840
aggttgaggt gggtagatta tttgagattg ggtattggag attagtttgg ttaatatgga      900
gaaattttgt ttttattaaa aatataaaat tagttggggtg tggttttgtaat                960
tatagttatt agggaggttg aggtaggaga attgtttgta tttgggaggt ggaggttgtg     1020
tgagttaaga ttgtgttatt gtattttagt ttgggtaatt agaataaaat tgttttaaaa     1080
aaaaaaaaaa taaaaatagg ttgggtgtgg tggtttatgt ttgtaatttt agtattttgg     1140
aaggttaagg tgggtggatt atgaggttag gagattgaaa ttattttggt taatatggtg     1200
aaattttgtt tttattaaaa aatataaaaa attagtgggg tgtggtggtg tgttgtag     1260
ttttagttat ttggggggtgt tgaggtagga gaatggtgag aatttgggag gtagagtttg     1320
tagtgagttg agattgtgtt attgtagttt ggtttgggtg aaagagtgag attttgtttt     1380
aaaaaaaaaa aaattaaata aataaataaa taaaaaaata aaataatagg agaagatgat     1440
gatagatttt ataagagttt ggtttttaggg ttgaaatttt atttagtgtt gtggtttttgt     1500
ggttgtgggt ttatgggtgt ttttaaaggg ttttaggtag atggaggttt ttggttggtg     1560
tggaggaagg aaaaggatga aattggtttt tttggttatt tttttttagtt tagttgtagt     1620
tgtagtttgt gggtttgagt tgattttttgg ggtttattgg tttattttat agatgtggag     1680
tttatagttg ggaggtgtgg tggttttggg tggtgtaggg ttttggggtg gggaggttga     1740
ggtgtttgtt ttttttgttg gggtttatttt gggttagttt tgagtttttg agtagggtgg     1800
attggttgtt gtttgtttaa agttggggtt gttaggagtg tgattgtttg tggatttgta     1860
gttgggtatt ggttgtgaatg tttgttttgt gatggtaggt gagatgtggg tggttttgtg     1920
ggaggggttt ttgtttgttt ttgttattta tttttttttttt gtttgttgtt gttgtagggt     1980
tgggaatggg ggggattaag tattttgtag ggtgatttgg ggagatttgt gagttagtgg     2040
ttttagtgtt gaatgagagt ttttttttttgg gtttgtgttt gtagggattg aagggattt     2100
```

```
tgaatgtttg tgggatgtgt tgtggtggtg tttagtatgg aggaggtagg tagggaggtg    2160
aggagtaggt tagattggtt gtaggtgtag ggttttttgt tttttaggt tgtggttttg      2220
tggagggttt gtagatagtt gttgagatta tgagggttag aggtttgtgg ttatgtgtgt     2280
tattattagt tagttgtttt ttgtttgg                                        2308
```

<210> 27
<211> 2553
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 27

```
ttggttattt ttgttttttt ttagattttt agagttttta gtgtagttat agaggatgtt    60
gttggtttta gttttaagaa gatgttgttt tttttagagg gttaagtaag tgggaatttt    120
ttttttatt tgttttgggt tttaggtagg gtttttggtg taaggtttgg ggttggaagt     180
tgatttattt aggtttaggt tttggggtag aattgaaatt ttttggttat tgttggttgt    240
agtttgggag taggttattg ttaaagttgt gggttttttt aggatagttt ttttatgagg    300
ttggtttttt atttgttgtt ttttatatt tggtggttag ggatgtggtt ttgggtagaa     360
tgatgatttt ttatttttgt tattatggaa gttattgttg tttttagtt tagttagtta     420
tttgggttgt agagtatttt ttttatgttt tttgggtgtt tttttttttt tttgttttag    480
tttggttttg ttttattttg ttttaggga ggggtatttt ggagtggggt tagggtatgg     540
ttttttttt agggagtttt ttttttggttg ttttagggt agtttttgtat agtttttagta   600
tttggtgtat ttttttttgat attttttta gggatagtta ggtatttttgt gtggggtatt   660
taagagagtt aggtttgtta gttttagtt tttgttagaa tgtaggtttg aggggtgagg     720
ggtggggtag gggtagggat aggaattttg gtgtgttttt tatttgtaaa ggtttattga    780
ggttttgagt tttagttatt gagttattaa gttagttgg gttaggtttg ggtgtttttgt    840
ttgtaatgga ggtagagatg gggttttggg gtagtttga ggatgttggg tgtatagtgg     900
gggttttgtt ggtaggaatt atttatgttt ttttttgggt taagtttgt ggatgtttag     960
tttgggggttg tggggagttg gtaggttagt ggtagatatt ggtgggtaga tttagtgttt   1020
ggtagaatag gtattaagga agtggtgatt ggagggaagt taagtgtatt taaatttttg    1080
ggtgagttat tattgttggg tttttttatag ttgttgaaag tgagtaatag tgatgaaggt   1140
ttgtgagttt ttgtgtgagt gagtgaatgg attagtagta gtttttaggt tgtggaagag    1200
tgttttttttt ttgggatggg gatatttggt tatagtaatt tttaattttt tatttatta    1260
ttgtttattg tagaggtatg tgggggtttt gtttttgta ggtaggagtg aggggtattt     1320
ttgtgatgtg gtatttttgt gattgaggtt atgtgtgatt ggtgtaaggg taggaagtga    1380
gttattggtt tgtattaggt gtggggtttt ttgtgagggt aggatttaaa gttggttgg    1440
ttttttggtt gtagtatttt tttttttttt gaattaggtt agagttttgg gatgggaggt    1500
gttttgtaga ttattttttt tattaatttt tgtttttttgt tttattttg tggtgatttg   1560
gtgaattgtt ggtttttttgt tgtgtattga gtggggtagt gatttttgatg tggtgttttt  1620
tgttgttttt gttattgtta ttatttttgg tggtttagtt tttgtatttt ttatttttat    1680
ggaggaatgt attaggtttt ttttttggga tgtatttttt atttatatgt ttttaaattt    1740
tggtattttt tgttttttttt ttattttat tttttttttt aggttttag ataaagggga     1800
agtggttgga tttttttaaa gggatagtgt tttattagtt tattgttgaa ttttttttttt   1860
taattttagt ttttagttta tagttaatta gtattagtag atagtttatg agtgatattt    1920
atgtaggttt taggttgtgg agagtttttt gggtaggaaa tagttttaa ggtttttttat    1980
tttatttagg ttttagtttt ttttatttgt ttttttttta gattgtggtt tttttggagtt   2040
tggttttttt gttttgtgt gattgatata tagtatttaa atagtggtag agtgggatgg      2100
atttttttagt ttgttttttg tgtgggtttg tattttgatt tagatatgtt ttttatagt    2160
aggatttagg ggggtatatg tgtgtttgtg ggttattgg ggtatttgta tttggtttat      2220
tttatttttt agagagaggg tttttgttgtg ttatttagtt ggagtgtagt ggtgtaatta   2280
tagtatattg tagttttttaa ttttttgggtt taagtgatt tttttttttta gttttttttag  2340
tagttgggag tataggattt attgtatttt ggttaatttt ttaataattt tttaagagat    2400
ggggttttat tgtgttgttt aggttggttt taaatttttg gttttaagtg attttttttat   2460
ttttgttttt tgaagtgttg agattatagg tatgagttat tatgtttatt ttagattgat    2520
atttttatat ttgtttattt tggttgggta ggg                                  2553
```

<210> 28
<211> 2553
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 28

```
ttttgtttag ttaggatgaa taaatataga aatgttagtt tgggatgggt atggtggttt    60
atgtttgtaa ttttagtatt ttaggaggtg aaggtgggag gattatttga ggttagaggt    120
```

```
ttgaggttag tttgggtaat atagtaagat tttatttttt aaaaaattat taaaaaatta      180
gttaggatat agtgggtttt gtatttttag ttattaggga ggttggggag ggaggattgt      240
ttgagtttag gagttgaagg ttgtagtgtg ttatgattat attattgtat tttagttggg      300
tgatatagta agattttttt tttaaaaaat aaaataaatt aaatgtgggt gttttagtga      360
atttgtaggt atatatgtgt ttttttgggt tttgttgtgg gggggtatgt ttgggttagg      420
gtatagattt atatagagaa taggttaggg ggtttgtttt gttttgttat tgtttgggtg      480
ttatgtgttg gttatatagg gatagaagaa ttaggtttta aagggttata atttaggaga      540
gaggtaggta ggaaagattg ggatttggat gagatgaggg attttaaggg ttgtttttta      600
tttaggaaat tttttatagt ttggggtttg tatgggtatt atttatgggt tgtttgttaa      660
tgttaattaa ttgtaattag ggaattgggg ttgaggaggg gagtttagta gtaagttggt      720
gggatattgt tttttaaga ggatttagtt attttttttt tgtttgggag tttaggggaa      780
ggggtgggag taaaggggga gtagaaaatg ttagggtttg gaagtatgtg ggtgaggggt      840
gtatttagga agggaggtgg ggtgtatttt tttatggggg tggggaatgt ggaggttggg      900
ttattggagg tggtggtggt ggtaggggtg gtaggagatg ttatgttagg gttattgttt      960
tatttggtgt atagtagggg gttggtagtt tattggatta ttgtgggggt ggggtggagg     1020
atggaagttg gtgagggggg tggtttatag ggtatttttt gttttagggt tttaatttaa     1080
tttgaaaggg aaaggagtgt tgtagttggg aggttaggtt gattttgggt tttgtttttg     1140
tagaagtttt tatgtttggt gtagattaat gatttgtttt ttgtttttat attgattata     1200
tatgattttg gttataggg tgttatatta taggagtgtt tttattttt gtttgtagga     1260
agtagggttt ttgtatattt ttgtagtggg tggtgggtgg gtgggggatt aggaattgtt     1320
gtaattaagt gtttttattt tggggaggga atgttttttt ataatttgga aattgttatt     1380
ggtttatttta tttgtttatg tagaaattta taaatttta ttattgttgt ttatttttag     1440
tagttgtgaa agatttggtg gtgatgattt atttgagggt ttgagtgtat ttggtttttt     1500
tttggttatt atttttttga tgtttgtttt attagatatt aggtttgttt attagtgttt     1560
gttattgatt tgttagtttt ttgtggtttt aggttgggta tttataaagt ttggtttagg     1620
agagagtata agtgattttt gttggtgagg ttttttgttgt gtatttagta tttttagaat     1680
tgttttgaga ttttgttttt gtttttattg tagatagggt atttaggttt ggtttaggtt     1740
gatttggtgg tttagtggtt ggggtttggg gttttagtga attttgtggg atggagagta     1800
tgttggagtt tttgtttttg tttttgtttt gttttttatt ttttaggttt gtattttggt     1860
aggagttaga ggttgatggg tttggttttt ttgagtgttt tatatagagt gtttgattgt     1920
ttttaagaag gatgttaagg gaggtgtgtt aggtattgag gttgtgtaga gttgtttttgg     1980
ggatagttag agaggaattt ttttggggga gagttatgtt ttggtttttat tttagggtat     2040
tttttttga gagtagggta ggataaagtt aggttggggt aggagaaggg ggaggtattt     2100
ggagggtatg aaaggggtgt tttgtagttt aggtggttgg ttgggttggg agatagtggt     2160
ggttttttata atgataggag tggagaatta ttgttttatt tagggttata ttttttggtta     2220
ttaggtgtga agaaatagta ggtggaggat tggttttatg gggagattgt tttggaagga     2280
tttatagttt tggtagtggt ttgttttttag gttgtagttg atagtaatta aggagtttta     2340
gttttgtttt agagtttgga tttaggtggg ttggtttttta gttttaggtt ttatattagg     2400
ggttttgttt ggagtttagg ataagtaggg agggggattt ttatttattt agttttttgg     2460
aggaagtggt gtttttttgg ggttgaagtt aatagtattt tttgtggtta tattgggggt     2520
tttggaggtt taaggagggg taggggtggt tag                                  2553
```

<210> 29
<211> 2308
<212> DNA
<213> Homo Sapiens

<400> 29

```
ccaggcagga ggcagctagc tggtgatagc gcgcgtgacc acagacctct gaccctcata       60
gtctcagcag ctgtctgcag gccctccacg ggaccgcagc ctgggagggc aggaggccct      120
gcacctgcgg ccggtctgac ctgctcctca cctccctgcc tgcctcctcc gtgctgggcg      180
ccgccgcgac acgtcccgcg gacattcagg agtcccttca gtccctgcga gcgcggaccc      240
gggagagagc tctcattcaa cgctgaggcc actggctcac agatctcccc gggtcaccct      300
gcgaagtgct tggtcccccc cattcccagc cctgcagcag caacaaacg gaaggaagtg      360
ggtgacgagg acaggcagag acccctccca cggggccgcc cgcatctcac ctgccgtcgc      420
ggggcgggcg ttgcggccgg tgcccagctg cagatccgca ggcgatcgca ctcctggcgg      480
ccccagcttt aagcaggcag cggccgatcc gccctgctcg agggctcagg gctgacccag      540
gtgagccccg gcggggaggg cgggcgcctc ggcctccccg ccccggaacc ctgcaccgcc      600
cagagccgcc acgcctcccg gctgtaggct ccgcatctgt aaaatgggcc gatgaacccc      660
aagggtcggc tcagacccac ggactgcagc tgcagctggg ctagaagggg tggccagggg      720
agccgatttc atccttttcc ttcctccacg ccaaccagga gcctccatct gcctgggacc      780
ctttgggggc acccatgggc ccacagccgc agaaccacag cactgggtgg ggtttcagcc      840
ctgggccaa gctcttgtga ggtctgtcat catcttctcc tgttgtttg ttttttgtt      900
tgtttgtttg tttggttttt tttttttga dacagagtct cgctctttca cccaggccgg      960
actgcagtgg cgcgatctcg gctcactgca agctctgcct cccgggttct cgccattctc     1020
ctgcctcagc accccgagt agctgggact acgggcgccg ccaccgcgcc cggctaattt     1080
tttgtatttt ttagtagaga cggggtttca ccgtgttagc caggatggtt cgatctcct     1140
gacctcgtga tccacccacc ttggccttcc aaagtgctgg attacaggc gtgagccacc     1200
gcgcccggcc tgttttgtt tttttttt ttgagacagt tttattctgg ttgcccaggc     1260
tggagtgcag tggcgcaatc ttggctcacg caacctccgc ctcccgggta caagcgattc     1320
```

```
tcctgcctca gcctccctag tagctgtgat tacaggcacg caccaccacg cccggctaat       1380
tttgtatttt tagtagagac ggggtttctc catgttggcc aggctggtct ccaatgcccg       1440
atctcaggtg atctgcccac ctcagcctcc caaagttctg cgattacagg cgtgagtcac       1500
catgcctggc cagtctcctg ttttttttaaa aacttattat ttagccggtg tggtgtctca      1560
cacctgtaat cctggcactt tgggaggccg aggtgagcgg atcacttgag gtcaggtgtt       1620
ccagaccagc ctgggcaaca tggcaaaacc ttgtctctac taaaaagaca aaaattagct       1680
ggaggtgctc tcttgaaccc aggaggcaga ggttgcagtg aactgagatc atgccactgc       1740
actccaaccc aggaggagga ggttgcactg agccaatatc atgccactac attccagcct       1800
gggcgacaga gcaagactct gtcccccca aaaaaaaaac agaaaaaaaa aaaagaaag        1860
aaatctgaag caattatgac caaatgttaa catctgtttc ttcacgtggt gggtacaggc       1920
gtgttttgga agcttctatg cttcctgtat tttgagatca cacaaaatta tgtttattat       1980
cttattgtgc tacatgatct ttgtagaaaa gttggtaaat tctggaatgc acgaagaaaa       2040
agattataat agccagtaat acccctgccc agagagacag tatccattaa tttttctctg       2100
cctatgtttg tcttcggaaa aatgggatac tattggcact atttttataac ttttttaatttt     2160
gtattaacac tttctatatt ttttaaattg aaaaagctgg ccaagcatgg tggctcaggc       2220
ctgtaaaccc agccaatgtg ggaggatcgc ttgagcccgg gagttcagga acagcctggg       2280
catcgtggcg agaccacatt gctacaaa                                         2308
```

<210> 30
<211> 2553
<212> DNA
<213> Homo Sapiens

<400> 30

```
ctggccaccc ctgccctccc ttagacctcc agagccccca gtgtagccac agaggatgct        60
gttggcttca gccccaagaa gacgccgctt cctccagagg gctaagtaag tgggaatccc       120
cctccctact tgtcctgggc tccaggcagg gccctggtg taaggcctgg ggctggaagc       180
cgacccacct aggtccaggc tctggggcag aactgaaact ccttggttac tgtcggctgc       240
agcctgggag caggccactg ccaaagctgt gggtccttcc aggacagtct ccccatgagg        300
ccggtcctcc acctgctgtt tcttcacacc tggtggccag ggatgtggcc ctgggtagaa       360
cgatgattct ccactcctgt cattatggaa gccaccgctg tctcccagcc cagccagcca      420
cctgggctgc agagcacccc tttcatgccc tccgggtgcc tccccttct cctgccccag        480
cctggctttg tcctaccctg ctctcaggga ggggtaccct ggagtggggc cagggcatgg       540
ctctccccg agggagttcc tctctggctg tccccagggc agctctgcac agcctcagta         600
cctggcgcac ctcccttgac atccttctta gggacagtca ggcactctgt gtggggcact      660
caagagagcc aggcccgtca gcctctagct cctgccagaa tgcaggcctg aggggtgagg      720
ggcggggcag gggcagggac aggaactccg gcgtgctctc catccgcaaa ggttcactga      780
ggccccgagc cccagccact gagccaccaa gtcagcctgg gccaggcctg ggtgccctgt       840
ctgcaatgga ggcagagacg gggtctcggg gcagttctga ggatgctggg tgcacagcgg      900
gggcctcgcc ggcaggaatc acttatgctc tctcctgggc caagctttgt ggatgcccag       960
cctgggggcg cggggagctg gcaggtcagt ggcagacact ggtgggcaga cctagtgtct      1020
ggtagaacag gcatcaagga agtggtgacc ggagggaagc caagtgcact caaaccctcg      1080
ggtgagtcat caccgccggg tctttcacag ctgctgaaag tgagcaacag tgatgaaggt      1140
ttgtgagttt ctgcgtgagc gagtgaatgg accagtagca gtttccaggt tgtggaagag      1200
cgttccctcc ccgggatggg gacacttggt tacagcaatt cctaatcccc cacccacccc      1260
ccgcccactg cagaggtatg cggggggcct gcttcctgca ggcaggagtg aggggcactc      1320
ctgtgatgtg gcacccctgt gaccgaggtc atgtgtgatc ggtgtaaggg caggaagcga      1380
gtcattggtc tgcaccaggc gtgggggctt ctgcgagggc aggacccaaa gtcggcctgg      1440
cctcccggct gcagcactcc tttccctttc gaattaggtt agagccctgg gacgggaggt      1500
gccctgtaga ccacccccct caccaacttc cgtcctccgc cccaccccg cggtgatccg       1560
gtgaactgcc ggccccctgc tgtgcaccga gtggggcagt gaccctgacg tggcgtctcc      1620
tgccgcccct gccaccgcca ccacctccgg tggcccagcc tccgcattcc ccacccccat      1680
ggaggaatgc accaggcctc ccttcctgga tgcaccctc acccacatgc ttccaaaccc        1740
tggcattttc tgctcccct ttactcccac cccttcccct aggctcccag acaaagggga        1800
agtggctgga tcctcttaaa gggacagtgt cccaccagct tactgctgaa ctcccctcct      1860
caaccccagt tccctagtta cagttaatta gcattagcag acagcccatg agtgatacce      1920
atgcaggccc caggctgtgg agagtttcct gggtaggaaa cagcccttaa ggtccctcat      1980
ctcatccagg tcccagtctt tcctacctgc ctctctccta gattgtggcc ctttggagcc      2040
tggttcttct gtccctgtgt gaccgacaca tagcacccaa acagtggcag agcgggacgg      2100
accccctagc ctgttctctg tgtgggtctg taccctgacc cagacatgcc cccccacagc      2160
aggacccagg ggggcacatg tgtgcctgcg ggttcactgg ggcacccgca tttggtttat      2220
tttatttttt agagagaggg tcttgctgtg tcacccagct ggagtgcagt ggtgtaatca      2280
tagcacactg cagccttcaa ctcctgggct caagcgatcc tccctcccca gcctccctag      2340
tagctgggag tacaggaccc actgtatcct ggctaatttt ttaataattt tttaagagat      2400
ggggtcttac tgtgttgccc aggctggcct caaacctctg gcctcaagtg atcctcccac      2460
cttcgcctcc tgaagtgctg agattacagg catgagccac catgcccatc ccagactgac      2520
atttctatat ttgttcatcc tggctgggca ggg                                    2553
```

<210> 31

<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 31

gggattattt ttataaggtt   20

<210> 32
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 32

ctctaaaccc catcccc   17

<210> 33
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 33

cccatccccca aaaacacaaa ccac   24

<210> 34
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 34

cgtcgtcgta gttttcgtt   19

<210> 35
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 35

tagtgagtac gcgcggtt   18

<210> 36
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 36

accgaaaata cgcttcacg   19

<210> 37
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 37

gcgttatcgt aaagtattgc gc   22

<210> 38
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 38

cgcgacgaac aaaacgccg   19

<210> 39
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 39

cgcgctactc cgcataca   18

<210> 40
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 40

gaggtaatcg aggcggtcg   19

<210> 41
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 41

cgccaattca tacgccgcac c   21

<210> 42
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 42

ttgtggttcg ggaagagac   19

<210> 43
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 43

tcccgaactc ttcgatcg   18

<210> 44
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 44

aactacgcgc aaacccgcga   20

<210> 45
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 45

cgcgctactc cgcataca   18

<210> 46
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 46

gaggtaatcg aggcggtcg   19

<210> 47
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 47

cgccaattca tacgccgcac c   21

<210> 48
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 48

accgaaaata cgcttcacg   19

<210> 49
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 49

gcgttatcgt aaagtattgc gc   22

<210> 50
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 50

cgcgacgaac aaaacgccg   19

<210> 51
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 51

gcgttttacg tcgtcgcg   18

<210> 52
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 52

gacgctaaac gccaccgt   18

<210> 53
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 53

ccgaccatcc gacgccttac tcg   23

<210> 54
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 54

cgttttttcgt tttattttcg c   21

<210> 55
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 55

gacaaaaaac gccacgtc   18

<210> 56
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 56

ccgacaattc accgaatcac cg   22

<210> 57
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 57

atctcaccta ccgtcgcg   18

<210> 58
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 58

taggagtgcg atcgtttgc   19

<210> 59
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 59

acgaacgtta cgaccgatac ccaacta   27